Heterogeneous information fusion: combination of multiple supervised and unsupervised classification methods based on belief functions

Na Li, Arnaud Martin, Rémi Estival

Heterogeneous information fusion: combination of multiple supervised and unsupervised classification methods based on belief functions

Na Li[a,b,*], Arnaud Martin[a], Rémi Estival[b]

[a] *Univ Rennes, CNRS, IRISA, DRUID, Rue E. Branly, 22300 Lannion, France*
[b] *TOTAL CSTJF, Avenue Larribau, 64000 Pau, France*

Abstract

In real-life machine learning applications, a common problem is that raw data (*e.g.* remote sensing data) is sometimes inaccessible due to confidentiality and privacy constrains of corporations, making classification methods arduous to work in the supervised context. Moreover, even though raw data is accessible, limited labeled samples can also seriously affect supervised methods. Recently, supervised and unsupervised classification (clustering) results related to specific applications are published by more and more organizations. Therefore, combination of supervised classification and clustering results has gained increasing attention to improve the accuracy of supervised predictions. Incorporating clustering results with supervised classifications at the output level can help to lessen the recline on information at the raw data level, so that is pertinent to improve the accuracy for the applications when raw data is inaccessible or training samples are limited.

We focus on the combination of multiple supervised classification and clustering results at the output level based on belief functions for three purposes: (1) to improve the accuracy of classification when raw data is inaccessible or training samples are highly limited; (2) to reduce uncertain and imprecise information in the supervised results; and (3) to study how supervised classification and clustering results affect the combination at the output level.

Our contributions consist of a transformation method to transfer heterogeneous information into the same frame, and an iterative fusion strategy to retain most of the trustful information in multiple supervised classification and clustering results.

Keywords: belief functions, heterogeneous information fusion, combination of classification and clustering

☆This document is the results of the research project funded by TOTAL (Ref: FR00008871).

*Corresponding author

Email addresses: lina05041992@gmail.com (Na Li), arnaud.martin@irisa.fr (Arnaud Martin), remi.estival@total.com (Rémi Estival)

1. Introduction

Ensemble methods can efficiently take advantage of information from different classifiers and thus outperform any single one [11]. However, in many real-life applications, raw data for machine learning methods is sometimes inaccessible due to confidentiality, and only results of classification and clustering are available. In addition, despite the possible accessibility of raw data, limited labeled samples can also induce low performance of supervised classifiers and ensemble methods.

Current ensemble methods can be classified into two aspects: the level of supervision (supervised, semi-supervised, unsupervised), and the level of execution of the ensemble process (no ensemble, ensemble at the raw data level and the ensemble at the output level) [15], [16]. Ensemble methods in each category have their specific advantages and limitations. Supervised ensemble methods can easily outperform any single classifer. Nevertheless, they rely heavily on training samples in classification problems [17]. In general, a long-standing rule of thumb in the supervised context requires training samples to be 10 times, preferably 100 times, more than the number of variables. Unfortunately, it is generally challenging to guarantee sufficient training samples in many complex situations. Conversely, unsupervised ensemble methods do not require training samples, yet generating unlabeled predictions [7], [36]. Semi-supervised ensemble methods overcome the constraints on training samples and thus has gain increasing attention in recent studies [3],[30].

Ensembles at raw data can fully exploit information of data. However, in numerous applications, raw data from different sources have different formats, making it difficult to combine them in the same scenario. More importantly, in real-world applications, raw data can sometimes be inaccessible due to confidentiality and privacy constrains of corporations, which makes ensembles arduous to work at the raw data level. Ensembles at the output level can reduce the dependence on the raw data, so that they are more pertinent to employ when limited information is available at the raw data level. Another benefit of the output level fusion is its higher robustness and less sensitivity on the type of data. It has also been proved that fusion at the output level is more powerful and effective in managing uncertainty and imprecision in complicated classification problems [26], [27].

Another important issue in ensemble methods is the management of uncertainty and imprecision that refer to epistemic states caused by imperfect or unknown information. Decreasing uncertainty can help to ameliorate predictions of ensemble methods. However, handling uncertainty at the output level is challenging for the current ensemble methods. We therefore propose a novel solution based on belief functions to reduce the uncertainty of supervised methods by combining with clustering results. Clustering results can provide information on compactness and separateness of data. Moreover, various numbers of clusters can reflect information on distributions of data, probably being different from distributions of classes. These information can not be reflected by classification results. Therefore, the combination with clustering results can provide heterogeneous yet supplementary information for classifications, so that it decreases the uncertainty caused by lack of knowledge or imperfect information in classification results.

In this paper, we propose an evidential fusion method to combine supervised and clustering results (EFSC) at the output level. The proposed EFSC can effectively combine information from clustering with classification and thus decrease the dependence on raw data and labeled samples. The uncertainty and imprecision at the raw data level can finally be reflected at the output level and therefore can be measured by belief functions. The proposed EFSC attempts to deal directly with uncertainty and imprecision at the output level of supervised results by combining several unsupervised results. In this way, the most reliable information from different sources can be extracted and then be combined to improve the overall accuracy.

The rest of the paper is organized as follows: In section 2 and section 3, we present the current semi-ensemble methods at output level and the basic concepts of belief functions. In sections 4, we propose a transformation method to transfer heterogeneous information into the same frame. In section 5, we propose an iterative fusion process to retain the most trustful information when combining multiple supervised and unsupervised predictions. A numerical example of EFSC is presented in section 6. In section 7, we compare the proposed method in framework of belief functions with current semi-ensemble methods on different data sets. In section 8, we conclude on our research work and present the perspectives.

2. Related works

Many previous efforts have shown the effectiveness of combining multiple machine learning methods. In this section, we briefly present several related studies, including ensemble methods at the raw data level in section 2.1 and at the output level in section 2.2.

2.1. Ensemble learning at the raw data level

Ensemble learning at raw data level can be also considered as learning in the supervised context, where the raw data is applied to train a combination process [32], [13]. Many studies have focused on the combination at the raw data level. For instance, multi-view learning, a popular method at raw data level, can learn from both labeled and unlabeled data from multiple sources [5], [37]. It aims to learn a function to model each view and jointly optimizes all the functions to improve the overall performance. Multi-view learning relies heavily on the initial performance of the supervised classifier. The significance of the initial accuracy of a basic classifier has been fully studied in [29], proving the supervised classifier is essential for the final combination results. In semi-supervised context, numerical studies have shown the remarkable performance of fuzzy methods [3],[30]. It has been experimentally proved that the low-fuzzy samples added to the training dataset can help to improve the accuracy of classification [30]. A novel fuzziness based semi-supervised learning approach is proposed in [3], which produces fuzzy memberships on unlabeled samples to assist with supervised methods. These two researches demonstrate the importance at the raw data level to use fuzziness, also a kind of uncertainty, to improve the accuracy of classification.

2.2. Ensemble learning at the output level

At the output level, major research has been devoted to ensemble methods for either classification or clustering. In the supervised context, majority voting [28] and belief functions [4], [25] are commonly used in combination at the output level. Compared to majority voting, the method based on belief functions can cope with more complex problems because several combination and decision rules are proposed in this framework to handle different situations [33], [22], [24]. In the unsupervised context, the combination is more difficult because single model has always different numbers of clusters. The final combination results are required to obtain the most agreement with each individual clustering, modeled by an objective function using consensus maximization [14]. Collaborative clustering [7], [36], is another family close to multi-view clustering, which allows different types of algorithms to work together. The theory of belief functions has also been used in combination with unsupervised methods in an early research [20], which has been proved to be effective for remote sensing data.

The combination of supervised and unsupervised methods has always been considered as a great challenge due to heterogeneous information in clustering. Less effective combinations have been proposed for high level fusion. Existing methods model the combination process as an optimization problem to find the most agreement with each individual method.

One of the early remarkable researches is the Bipartite Graph-based Consensus Maximization (BGCM) algorithm [15]. It considers results of clustering as constraints and maximizes the consensus between supervised and unsupervised predictions. C3E [1] is another early ensemble model that combine heterogeneous information. It uses multiple classifiers to generate an initial probability distribution at the class level for each object. Recently, the UPE model [2] has been proposed, casting the combination problem as an unconstrained probabilistic embedding problem. It assumes that objects and groups (classes/clusters) have latent coordinates without constraints in a D-dimensional Euclidean space. The prediction of an object is then determined by the distance between the object and the classes in the embedded space. A novel method, EC3 [6], has recently been proposed to merge classification and clustering by mapping the combination into a convex optimization problem. The objective function is based on the consensus at the object level as well as at the group level, which surpasses previous methods on different data sets.

Most of the existing methods considered clustering predictions as supplementary constraints for supervised results, consequently relying more on supervised methods than unsupervised ones. BGCM, C3E and EC3 take the probability of class distribution as the core of the objective function in which clustering results are used as supplementary constrains. It indicates that sufficient training samples are needed to ensure the performance of supervised methods, and on the other hand, unsupervised results have not been fully exploited. Less attention has been paid to the management of uncertainty and imprecision in the fusion process, which however includes a wealth of information to improve accuracy.

Belief functions, an effective theory for addressing uncertain and imprecise information, are usually considered to model inperfect information in combination. Numerous studies [31], [21], [23] have demonstrated the efficiency of belief functions in both classification and clustering. A novel method for combining supervised and unsupervised methods was proposed in [19], in which the results of clustering are modeled as discounting coefficients to generate mass functions in the same frame as classification. The experiments in [19] demonstrate a great potential for solving the combination problem by belief functions.

3. Basic concepts of belief functions

The theory of belief functions is a popular method for dealing with uncertainty and imprecision with the belief reasoning framework [34], [8]. This section presents the basic concepts of belief functions, which will be used in this paper.

3.1. Representation of information

Let's consider a decision from the source E regarding a variable X. All possible states ω_x of X construct a finite set Ω, called the frame of discernment. The information supporting the decision on X can be quantified by a basic belief assignment (BBA), also called mass function m^Ω, which projects 2^Ω on $[0, 1]$, verifying:

$$\sum_{A \subseteq \Omega} m_E^\Omega(A) = 1 \tag{1}$$

Function m_E^Ω represents the state of knowledge of the source E about the variable X on Ω. When there is no ambiguity, it can be simplified as m. We give the basic and important concepts of BBA as follows:

Definition 1 (*Focal element*). Focal element is a subset A of Ω verifying $m(A) > 0$.

Definition 2 (*Normal BBA*). A BBA is normal if $\emptyset$ is not a focal element.

Definition 3 (*Bayesian BBA*). A BBA is Bayesian when all its focal elements are singletons.

Definition 4 (*Simple BBA (SBBA)*). A BBA is simple if it has no more than two focal elements, Ω being included. In this paper, we denote the SBBA as m_A, where A is the focal element besides Ω.

Definition 5 (*Categorical BBA (CBBA)*). A BBA has only one focal element A, denoted as $m_{[A]}$. We have $m_{[A]}(A) = 1$.

Definition 6 (*Non-dogmatic BBA*). A BBA is non-dogmatic if Ω is a focal element.

In belief functions, uncertainty can be represented by BBA values on singletons, and imprecision refers to BBA values on unions. Specifically, the BBA value on Ω (*i.e.* $m(\Omega)$) is called the ignorance.

3.2. Discounting

Discounting operation is used to model the reliability of the source regarding a piece of information, and defined by:

$$^{\alpha}m(A) = \alpha m(A) + (1-\alpha)m(\Omega), \forall A \subseteq \Omega \tag{2}$$

with $\alpha \in [0,1]$. A discounting coefficient α reflects the reliability of the source. $\alpha = 1$ means the source is completely reliable and the information it provides can thus be entirely taken into account. On the contrary, a null α indicates that the source is not reliable at all and thus its information cannot be considered.

3.3. Least Committed Principle (LCP)

The Least Commitment Principle plays a significant role in belief functions, as does the principle of maximum entropy in Bayesian Probability Theory. Given several BBAs compatible with a set of constraints, the least informative one should be selected [35]. Many partial orderings are proposed on the set of belief functions to compare information by LCP. We can compare separable BBAs by the weight of evidence w, defined as:

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}, \forall A \subset \Omega \tag{3}$$

where q represents the commonality function, defined as:

$$q(A) = \sum_{B \supseteq A} m(B) \tag{4}$$

For a SBBA m_A, the weight of evidence can be simplified as $w(A) = -ln(m_A(\Omega))$. Given two non dogmatic BBAs m_1 and m_2, m_1 is considered more committed than m_2 if it verifies $w_1(A) \leq w_2(A), \forall A \subset \Omega$. This is noted as $m_1 \sqsubseteq_w m_2$ [10] called $w-$ordering, and we can also say m_1 is w-more committed than m_2.

3.4. A distance model to estimate BBAs

Distance BBAs estimation is suggested by Denoeux [9]. For a group B ($B \subseteq \Theta$) consisting of a set of objects, the closer to the center of B, noted as $\bar{B}$ object x is , the more certain in B x could be. This information can be modeled by a SBBA for the object x, $\forall B \subseteq \Theta$:

$$\begin{cases} m^{\Theta}(B) = \alpha e^{-\gamma d(x,\bar{B})}, \\ m^{\Theta}(\Theta) = 1 - \alpha e^{-\gamma d(x,\bar{B})}. \end{cases} \tag{5}$$

where α is a discounting coefficient and γ can be used to handle the lack of knowledge, and d represents the Euclidean distance.

3.5. Transformation of BBAs

A transformation proposed by Karem [19] is to project a BBA in Θ to Ω based on the similarity measured by proportion. For a group pf objects X, the similarity between A and B is modeled by a BBA m_{ns} verifying:

$$m_{ns}(A) = \frac{|\{x : x \in A\} \cap \{x : x \in B\}|}{|\{x : x \in B\}|}, A \subseteq \Omega, B \subseteq \Theta, x \in X \tag{6}$$

The transformation of a BBA from Θ to Ω, $\forall A \subseteq \Omega, \forall B \subseteq \Theta$, is defined by:

$$m^{\Theta\uparrow\Omega}(A) = m^{\Theta}(B) m_{ns}^{\Omega}(A) \tag{7}$$

$$m^{\Theta\uparrow\Omega}(\Omega) = 1 - m^{\Theta}(B)(1 - m_{ns}^{\Omega}(A)) \tag{8}$$

where m^{Θ} is the distance model presented in section 3.4

3.6. Combination of information

Two BBAs m_1 and m_2 representing information from two independent and reliable sources, can be combined by the conjunctive rule. For all $A \subseteq \Omega$, we have:

$$m_1 \text{ⓒ} m_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \tag{9}$$

Dempster's rule is equivalent to the normalized conjunctive rule, assuming $m_1 \text{ⓒ} m_2(\emptyset) \neq 1$, using equation (10), defined for all $A \subseteq \Omega$ by:

$$m_1 \oplus m_2(A) = \begin{cases} \frac{m_1 \text{ⓒ} m_2(A)}{1 - m_1 \text{ⓒ} m_2(\emptyset)} & \text{if } A \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

Both rules are commutative and associative and assume that combined evidence is cognitively independent and reliable.

If we only know at least one source is reliable, we can use the disjunctive rule. For all $A \subseteq \Omega$, we have:

$$m_1 \text{ⓓ} m_2(A) = \sum_{B \cup C = A} m_1(B) m_2(C) \tag{11}$$

3.7. Canonical decomposition

Canonical Decomposition, proposed by Shafer [34], describes that a BBA can be considered as the results of $\oplus$ combination of a group of SBBAs as follows:

$$m = \bigoplus_{\emptyset \neq A \subset \Omega} m_A \tag{12}$$

If the unique decomposition exists, m is said separable.

3.8. Decision based on BBAs

Decision is the last step in the framework of belief functions, based on combined mass functions from multiple sources. For each object, a subset of the frame of discernment has to be chosen to maximize a certain criterion. In belief functions, several decision rules are avaiable. We use the minimum Jousselme distance [12] to make decisions.

The minimum Jousselme distance (the minimum loss of confidence):

$$x \in A, \quad \text{if } d_J(m(x), m_{[A]}) = \min_{A \in 2^\Omega}\{d_J(m(x), m_{[A]})\} \tag{13}$$

where $m(x)$ is the BBA of object x, and $d_J(m(x), m_{[A]})$ is Jousselme distance between the $m(x)$ and the categorical $m_{[A]}$ with A as the focal element. Jousselme distance [18] between two BBA m_1 and m_2 is defined by:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T J (m_1 - m_2)} \tag{14}$$

where J is the Jaccard weighting matrix defined as:

$$J = \frac{|A \cap B|}{|A \cup B|}, \forall A \subseteq \Omega, \forall B \subseteq \Omega \tag{15}$$

4. Transformation of heterogeneous information in belief functions

In this section, we present the proposed transformation method in the framework of belief functions. This method can transform both uncertainty and imprecision in a BBA from its original frame of discernment to any other target frame, based on the similarity at the output level. Section 4.1 explains how we measure and model the similarity information between classification or clustering results in different frames of discernment. We illustrate the proposed transformation for heterogeneous information in section 4.2.

4.1. Similarity information

Jaccard Index originally is used to measure similarity between two finite sample sets. In land cover classification problem, a class and a cluster can be considered as two different finite sets which may contain some pixels in common. Therefore, we can use Jaccard Index to measure similarity between a class and a cluster. The more pixels they have in common, the more similar they are. For a group of objects X, suppose we have two methods c and s such that c can separate X in frame of discernment Θ and s in Ω. Here c and s can be a classification or a clustering. Jaccard Index is defined based on the classification/clustering results as:

$$Jac(T_i, O_j) = \frac{|\{x : x \in O_j\} \cap \{x : x \in T_i\}|}{|\{x : x \in O_j\} \cup \{x : x \in T_i\}|}, O_j \subseteq \Omega, T_i \subseteq \Theta, x \in X \tag{16}$$

A matrix $JM = [Jac(T_i, O_j)]$ can be used to measure the similarity of the classification/clustering results about any x in Ω and Θ. Apparently, the element $Jac(T_i, O_j)$ can be only affected by the subsets it involves, so it is independent of the other elements in the matrix. $Jac(T_i, O_j)$ gives uncertainty on the information that T_i and O_j are the same and nothing more, which can be modeled by a SBBA $m^{H_{ij}}$ in the frame of discernment $H_{ij} = \{Y_{ij}, N_{ij}\}$. Y_{ij} represents the class O_j and the cluster T_i are the same while N_{ij} indicates they are different. The information from a Jaccard Index between O_j and T_i can be modeled as a SBBA by:

$$\begin{cases} m_{ij}^{H_{ij}}(Y_{ij}) = Jac(T_i, O_j), \\ m_{ij}^{H_{ij}}(H_{ij}) = 1 - Jac(T_i, O_j). \end{cases} \tag{17}$$

where $m_{ij}^{H_{ij}}(Y_{ij})$ indicates the support on the information that O_j and T_i are the same, and $m_{ij}^{H_{ij}}(H_{ij})$ represents we know nothing about the relationship between O_j and T_i.

4.2. The proposed transformation of BBA

The information provided by the method c in the frame of discernment Θ can be modeled by a BBA m_c^{Θ}. To combine it with BBAs in the frame of discernment Ω, the crucial step is to transfer m_c^{Θ} as m_c^{Ω}. The transformation proposed in [19] takes only uncertainty on singletons from clustering, yet ignoring imprecision. To fully exploit the information in clustering, we decompose the original BBA m_c^{Θ} into a set of SBBAs, so that both uncertainty and imprecision can be well preserved during the transformation. The proposed transformation consists of the following four steps:

Step 1: Decompose a BBA on the original frame of discernment to SBBAs

Suppose m_c^{Θ} is a separable mass which accordingly can be decomposed as several SBBAs as

$$m_c^{\Theta} = \bigoplus_{\emptyset \neq T_i \subset \Theta} m_{c,T_i}^{\Theta} \tag{18}$$

where $m_{c,\ T_i}^{\Theta}$ indicates the SBBA from the method c and its focal element besides Θ is T_i. We can denote $m_{c,\ T_i}^{\Theta}$ as:

$$\begin{cases} m_{c,\ T_i}^{\Theta}(T_i) = t, \\ m_{c,\ T_i}^{\Theta}(\Theta) = 1 - t. \end{cases} \tag{19}$$

The SBBAs after decomposition can be considered as multiple independent sources that entirely preserve the information from m_c^{Θ}.

Step 2: Transfer a SBBA on the original frame of discernment to the target frame of discernment

We denote the transferred SBBA on T_i to O_j as m_{T_i,O_j}^{Ω}. It indicates the uncertainty about the information that object x clustered as T_i should be labeled as O_j, incorporating the similarity between T_i and O_j.

To calculate $m^{\Omega}_{T_i,O_j}$ in Ω, we can use the uncertainty on T_i or Θ as a discounting coefficient to modify the similarity BBA $m^{H_{ij}}_{ij}$. There are two strategies for performing this simple transformation: <1> using the uncertainty on T_i as a discounting coefficient to weaken $m^{H_{ij}}_{ij}(Y_{ij})$; <2> using the ignorance on Θ as a discounting coefficient to weaken $m^{H_{ij}}_{ij}(H_{ij})$.

In strategy <1>, $\forall T_i \subset \Theta, \forall O_j \subset \Omega$, we have:

$$\begin{aligned} m^{1,\Omega}_{T_i,O_j}(O_j) &= m^{H_{ij}}_{ij}(Y_{ij}) * m^{\Theta}_{c,T_i}(T_i) \\ &= Jac(T_i, O_j) * t \end{aligned} \tag{20}$$

$$\begin{aligned} m^{1,\Omega}_{T_i,O_j}(\Omega) &= 1 - m^{H_{ij}}_{ij}(Y_{ij}) * m^{\Theta}_{c,T_i}(\Theta) \\ &= 1 - t * Jac(T_i, O_j) \end{aligned} \tag{21}$$

In strategy <2>, $\forall T_i \subset \Theta, \forall O_j \subset \Omega$, we have:

$$\begin{aligned} m^{2,\Omega}_{T_i,O_j}(O_j) &= 1 - m^{H_{ij}}_{ij}(H_{ij}) * m^{\Theta}_{c,T_i}(\Theta) \\ &= 1 - (1-t) * (1 - Jac(T_i, O_j)) \\ &= t + Jac(T_i, O_j) - t * Jac(T_i, O_j) \end{aligned} \tag{22}$$

$$\begin{aligned} m^{2,\Omega}_{T_i,O_j}(\Omega) &= m^{H_{ij}}_{ij}(H_{ij}) * m^{\Theta}_{c,T_i}(T_i) \\ &= (1-t) * (1 - Jac(T_i, O_j)) \\ &= 1 - t - Jac(T_i, O_j) + t * Jac(T_i, O_j) \end{aligned} \tag{23}$$

According to the Least Committed principle (LCP), we should choose the less committed transferred SBBA between strategy <1> and strategy <2>.

Lemma 1. For two give non dogmatic SBBAs $m^{1,\Omega}_{T_i,O_j}$ and $m^{2,\Omega}_{T_i,O_j}$, $m^{1,\Omega}_{T_i,O_j}$ is less committed than $m^{2,\Omega}_{T_i,O_j}$ with w-ordering.

Proof. According to equations (20), (21) (22), (23), $\forall t, Jac(T_i, O_j) \in [0, 1]$, we have:

$$m^{1,\Omega}_{T_i,O_j}(O_j) + m^{1,\Omega}_{T_i,O_j}(\Omega) = 1 \tag{24}$$

$$m^{2,\Omega}_{T_i,O_j}(O_j) + m^{2,\Omega}_{T_i,O_j}(\Omega) = 1 \tag{25}$$

For all $O_j \subset \Omega$, we always have:

$$
\begin{aligned}
2 &\leq \frac{1}{t} + \frac{1}{Jac(T_i, O_j)} \\
&\Leftrightarrow 2t * Jac(T_i, O_j) \leq t + Jac(T_i, O_j) \\
&\Leftrightarrow t * Jac(T_i, O_j) \leq t + Jac(T_i, O_j) - t * Jac(T_i, O_j) \\
&\Leftrightarrow m^{1,\Omega}_{T_i,O_j}(O_j) \leq m^{2,\Omega}_{T_i,O_j}(O_j) \\
&\Leftrightarrow -ln(m^{1,\Omega}_{T_i,O_j}(\Omega)) \geq -ln(m^{2,\Omega}_{T_i,O_j}(\Omega)) \\
&\Rightarrow w^{1,\Omega}_{T_i,O_j} \geq w^{2,\Omega}_{T_i,O_j} \\
&\Rightarrow m^{2,\Omega}_{T_i,O_j} \sqsubseteq_w m^{1,\Omega}_{T_i,O_j}
\end{aligned}
\tag{26}
$$

$m^{1,\Omega}_{T_i,O_j}$ is less committed than $m^{2,\Omega}_{T_i,O_j}$ and should therefore be selected, simplified as $m^{\Omega}_{T_i,O_j}$.

Step 3: Combine all the evidence from transferred SBBAs

The transformed SBBA $m^{\Omega}_{T_i,O_j}$ represents only one piece of evidence on object x in O_j. From the perspective of clustering, x can also be clustered into different singletons or unions in Θ. The imprecision should also be considered when reconstructing the supports for the assertion that x is in O_j. Since all SBBAs transferred from T_i ($\forall T_i \subset \Theta$) to O_j are independent, we can combine them by Dempster's rule:

$$
m^{\Omega}_{c,O_j} = \bigoplus_{\forall T_i \subset \Theta} m^{\Omega}_{T_i,O_j}, \tag{27}
$$

The SBBA m^{Ω}_{c,O_j} on Ω represents the evidence that object x is labeled as O_j knowing all the evidence, provided by the method c on Θ, that x belongs to each T_i ($\forall T_i \subset \Theta$).

Step 4: Combine SBBAs on the target frame of discernment

Since m^{Ω}_{c,O_j} represents only a piece of evidence on the focal element O_j, we have to combine all the evidence on Ω to obtain a normal BBA as:

$$
m^{\Omega}_{c} = \bigoplus_{\forall O_j \subset \Omega} m^{\Omega}_{c,O_j}, \tag{28}
$$

m^{Ω}_{c} indicates a normal BBA on Ω with the results of the method c which are initially generated on Θ.

Figure 1 illustrates the process on the proposed transformation method. m^{Θ}_{c} is first decomposed into a set of SBBAs, *e.g.*, $m^{\Theta}_{c,T_1}, ..., m^{\Theta}_{c,T_n}$. For each SBBA such as m^{Θ}_{c,T_1}, it can be transferred as a group of SBBAs on Ω using similarity information. By doing so, each focal element on Ω has a group of evidence from Θ, which can be combined by Dempster's rule. To obtain a normal BBA, the SBBA on Ω can also be combined by Dempster's rule. The proposed transformation is more cautious because both uncertainty and imprecision on Θ are preserved and transferred into Ω, whereas the transformation in [19] takes only into account uncertainty.

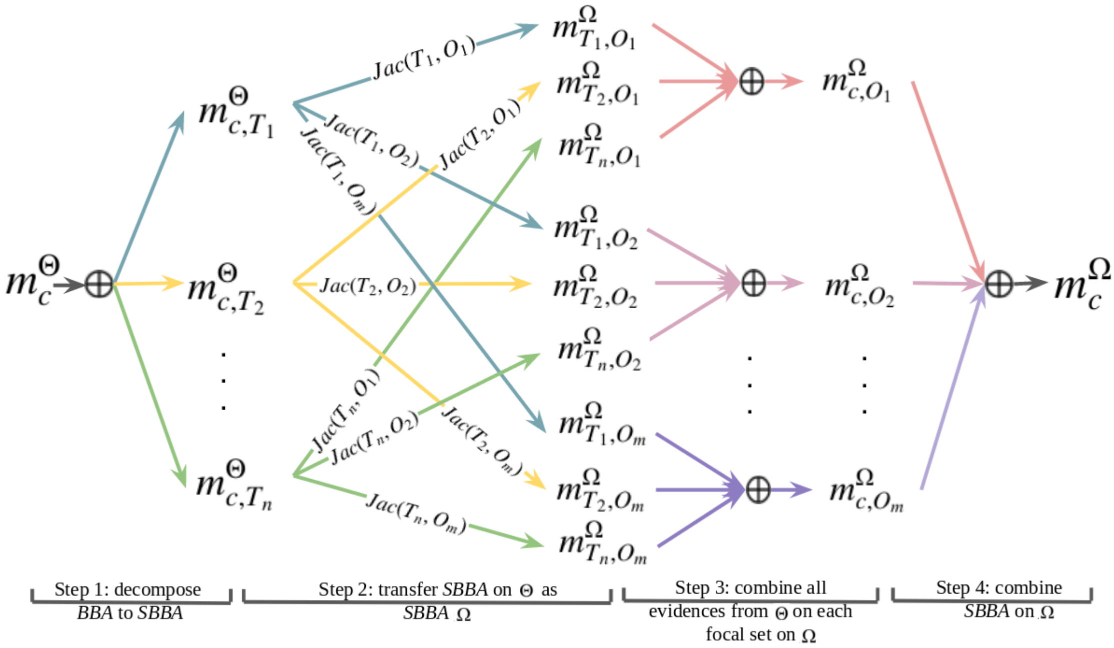


Figure 1: Transformation method to change the BBA in frame of discernment Θ to Ω.

5. Evidential fusion of supervised methods and clustering (EFSC)

The proposed transformation allows to transfer heterogeneous information from clustering to the same frame of discernment as classification, on the basis of which combining different classification and clustering methods becomes possible. Clustering with different numbers of clusters affects differently on the fusion with classification. To efficiently synthesize the information derived from clustering, we propose an iterative fusion process that combines several classification and clustering methods.

For a group of objects $X = \{x_1, x_2, ..., x_N\}$ to be classified, suppose we have a group of classification $S = \{s_1, s_2, ..., s_M\}$ and a group of clustering $C = \{c_1, c_2, ..., c_L\}$ with different numbers of clusters. We denote the frame of discernment of a classification as $\Omega = \{\omega_1, \omega_2, ..., \omega_x\}$ and that of a clustering as $\Theta = \{\theta_1, \theta_2, ..., \theta_y\}$. To simplify the problem, we can merely measure the similarity between each class ω_j and each cluster θ_i. Thus the similarity matrix JM can be simplified to the size $|\Omega| \times |\Theta|$.

5.1. Selection of partial information in clustering

We propose a criterion to control the iterative process to ensure that only useful information from clustering results is fused. In this case, we are more interested in the partial information about clusters rather than evaluate the whole clustering. Therefore, at each iteration we have to decide whether the information from certain clusters should be fused with the classes.

Let's denote, $\forall \omega_j \in \Omega$, $W_j^k = \{x : x \in \omega_j^k\}$, for all the objects classified as ω_j in iteration k. After combination with a clustering c_l, a new set of objects labeled as ω_j is generated, denoted as $W_{j,c_l} = \{x : x \in \omega_j\}$. This criterion is used to decide whether the information in W_{j,c_l} should be updated to W_j^k. Results updated by the criterion are denoted by W_j^{k+1}. Note that, if the information in W_{j,c_l} is synthesized, W_j^{k+1} becomes the union of W_j^k and W_{j,c_l}. When the uncertainty in W_{j,c_l} is less than that in W_j^k, we update the information in W_j^k by W_{j,c_l}. For each object $x \in W_j^k$, its uncertainty represented by BBA after decision is transformed to a scalar $R_{W_j^k}$, called by loss of confidence, defined based on Jousselme distance by:

$$R_{W_j^k} = d_J(m(x), m_{[\omega_j]}) \tag{29}$$

where $m(x)$ is the BBA of the object x and $m_{[\omega_j]}$ is the categorical BBA such that $m(\omega_j) = 1$. A smaller distance represents less loss of confidence on ω_j for object x. The loss of confidence of class ω_j, denoted $\overline{R_{W_j^k}}$, is the average loss of confidence of all objects x in W_j^k, as follows:

$$\overline{R_{W_j^k}} = \frac{\sum\limits_{x \in W_j^k} R_{W_j^k}}{|W_j^k|} \tag{30}$$

We update the information, *i.e.* labels, BBAs and losses of confidence in the iteration k of all objects in each class W_j^k when $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$. In some situations, due to insufficient information on certain classes, it is possible that no decision is made on these classes after combination. We therefore stop merging new information with these classes so that they can be retained during the iterative fusion process and has a relatively high uncertainty. Details of the criterion are given in the algorithm 1.

Note that due to the cautiousness of this criterion, the loss of confidence only decrease for classes that have less disagreement between classification and clustering. Let's denote $V_{j,c_l} = W_j^k \cap W_{j,c_l}$, $P_{j,c_l} = W_j^k - V_{j,c_l}$, $Q_{j,c_l} = W_{j,c_l} - V_{j,c_l}$. V_{j,c_l} represents a group of objects labeled as ω_j before and after fusion with clustering c_l. P_{j,c_l} indicates a group of objects, originally labeled as ω_j, yet which are not reclassified as ω_j after combining with clustering c_l. Q_{j,c_l} indicates a group of objects, not labeled as ω_j, but reclassified as ω_j after combining with clustering c_l.

For each $x \in P$, its information will not be changed by W_{j,c_l}. Only for each $x \in V_{j,c_l} \cup Q_{j,c_l}$, its information will be modified if the criterion is satisfied. The updated information may result in an increase in the average loss of confidence on W_j^{k+1} compared to W_j^k, even if it satisfies $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$, demonstrated as follows:

Lemma 2. For each $\omega_j \in \Omega$, given $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$, the average loss of confidence will be increased after the update, *i.e.* $\overline{R_{W_j^k}} < \overline{R_{W_j^{k+1}}}$ iff $\overline{R_{W_j^k}} < \overline{R_{Q_{j,c_l}}}$.

Proof. We have $\overline{R_{W_j^{k+1}}} = \overline{R_{W_j^k \cup W_{j,c_l}}}$ after the update.

$$\begin{aligned}
&\overline{R_{W_j^k}} < \overline{R_{W_j^{k+1}}} \\
&\qquad\qquad\Leftrightarrow \overline{R_{W_j^k}} < \overline{R_{W_j^k \cup W_{j,c_l}}} \\
&\qquad\qquad\Leftrightarrow \frac{|P_j|\overline{R_{P_j}} + |V_{j,c_l}|\overline{R_{V_{j,c_l}}}}{|P_j| + |V_{j,c_l}|} < \frac{|P_j|\overline{R_{P_j}} + |V_{j,c_l}|\overline{R_{V_{j,c_l}}} + |Q_{j,c_l}|\overline{R_{Q_{j,c_l}}}}{|P_j| + |V_{j,c_l}| + |Q_{j,c_l}|} \\
&\qquad\qquad\Leftrightarrow |P_j||Q_{j,c_l}|\overline{R_{P_j}} + |V_{j,c_l}||Q_{j,c_l}|\overline{R_{V_{j,c_l}}} < |P_j||Q_{j,c_l}|\overline{R_{Q_{j,c_l}}} + |V_{j,c_l}||Q_{j,c_l}|\overline{R_{Q_{j,c_l}}} \\
&\qquad\qquad\Leftrightarrow \frac{|P_j|\overline{R_{P_j}} + |V_{j,c_l}|\overline{R_{V_{j,c_l}}}}{|P_j| + |V_{j,c_l}|} < \overline{R_{Q_{j,c_l}}} \\
&\qquad\qquad\Leftrightarrow \overline{R_{W_j^k}} < \overline{R_{Q_{j,c_l}}}
\end{aligned} \tag{31}$$

When satisfying $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$, we must have $\overline{R_{V_{j,c_l}}} < \overline{R_{Q_{j,c_l}}}$ so that the average loss of confidence still increase after the update. However, the relationship between $\overline{R_{V_{j,c_l}}}$ and $\overline{R_{Q_{j,c_l}}}$ is not limited. Obviously, if we have $\overline{R_{W_j^k}} > \overline{R_{Q_{j,c_l}}}$, the average loss of confidence of class ω_j is a non-monotonic increasing function and can converge because it has the infinimum.

Algorithm 1: Criterion for partial information fusion

Input:

Supervised labels and BBAs on the frame of discernment of objects in a dataset $X = \{x_1, ..., x_N\}$.

Output:

Updated labels and BBAs on the dataset $X = \{x_1, ..., x_N\}$.

for *iteration* k **do**
  for $\forall \omega_j \in \Omega$ **do**
    if W_{j,c_l} *is not empty* **then**
      Calculate $\overline{R_{W_j^k}}$, the average loss of confidence on W_j^k in iteration k by equation (30).
      Calculate $\overline{R_{W_{j,c_l}}}$, the average loss of confidence on W_{j,c_l} by equation (30).
      if $\overline{R_{W_{j,c_l}}} < \overline{R_{W_j^k}}$ **then**
        Add information in W_{j,c_l} to W_j^k to obtain W_j^{k+1} :
        $m_x^{k+1} = m_x^{W_{j,c_l}}, \forall x \in W_{j,c_l}$
        $L_x^{k+1} = L_x^{W_{j,c_l}}, \forall x \in W_{j,c_l}$
        $R_x^{k+1} = R_x^{W_{j,c_l}}, \forall x \in W_{j,c_l}$
        $m_x^{k+1} = m_x^k, \forall x \in W_j^k - W_j^k \cap W_{j,c_l}$
        $L_x^{k+1} = L_x^k, \forall x \in W_j^k - W_j^k \cap W_{j,c_l}$
        $R_x^{k+1} = R_x^k, \forall x \in W_j^k - W_j^k \cap W_{j,c_l}$
      else
        Still keep the BBAs m^k, labels L^k, and the loss of confidence R^k in iteration $k+1$.
        $m_x^{k+1} = m_x^k, \forall x \in W_j^k$
        $L_x^{k+1} = L_x^k, \forall x \in W_j^k$
        $R_x^{k+1} = R_x^k, \forall x \in W_j^k$
    else
      Still keep the BBAs m^k, labels L^k, and the loss of confidence R^k in iteration $k+1$.
      $m_x^{k+1} = m_x^k, \forall x \in W_j^k$
      $L_x^{k+1} = L_x^k, \forall x \in W_j^k$
      $R_x^{k+1} = R_x^k, \forall x \in W_j^k$
  $k = k + 1$

5.2. Iterative fusion process

We repeat the previous step several times to reduce the uncertainty of the supervised results as much as possible until the overall loss of confidence is convergent. At each iteration, the information from a clustering is combined with the previous clustering by the Dempster's rule, thus strengthening the mass values on singletons and reducing the ignorance. As the combination accumulates, the results of the fusion become more and more certain on singletons, so the accuracy can be converged for a given pool of clustering methods C. The details of the iterative process are outlined in algorithm 2.

Algorithm 2: Iterative fusion process for one supervised method and multiple unsupervised methods

Input:

Labels generated from a supervised method c_m on test data $X = \{x_1, ..., x_N\}$: $L_{s_m}(x_1), ...L_{s_m}(x_N)$

Clustering results from a group of unsupervised methods $C = \{c_1, ...c_L\}$ on X: $[Q_{c_1}(x_1), ...Q_{c_1}(x_N)]$, ..., $[Q_{c_L}(x_1), ...Q_{c_L}(x_N)]$

Output:

Labels after combination of the supervised method s_m and multiple clustering methods in C

Loss of confidence based on Jousselme distance on X after fusion: R_{x_1},...,R_{x_N}

```
Calculate the BBAs of the supervised method s_m in the frame of discernment Ω.
Begin with the iteration step k = 0.
Calculate the average loss of confidence R̄^k in step k = 0, and initialize the average loss of confidence
  in step k = 1 as R̄^{k+1} = 0.
while |R̄^k − R̄^{k+1}| > ε do
    Randomly select a clustering c_l in C.
    Calculate the SBBAs for unsupervised c_l in its original frame of discernment Θ_l.
    Calculate the similarity matrix JM_ml of L_{s_m} and Q_{c_l} by equation (16).
    for ∀x ∈ X do
        Transfer its BBA of c_l in the frame of discernment Θ_l to Ω as the process shown in Figure 1.
        Update the information according to algorithm 1.
    Update R̄^k and R̄^{k+1}.
    k = k+1.
```

The proposed EFSC fusion strategy includes two principal steps: (1) enchaining the reliable information in each individual supervised method by randomly combining multiple unsupervised methods from C; (2) fusing the reliable information from multiple supervised methods.

After combination with a clustering, the updated fusion results are used as the supervised results to calculate the similarity with a new clustering at the next iteration. In this way, the similarity can be updated over the iterations and the information in the pool of clustering methods can be fully exploited when the iteration steps are sufficient. In algorithm 2, ε is a user-defined value close to 0 to control the stop condition.

The general workflow is detailed in Figure 2, where the different initial classifiers are denoted by s_1, s_2,..s_M. For the classifier s_1, for example, we have chosen at random z different clustering methods in C, denoted successively by $c_{s_{11}}$, $c_{s_{12}}$,...$c_{s_{1z}}$, to be combined with s_1 on the basis of the proposed transformation. In the i_{th} fusion step, new labeled information $L_{s_{1i}}$ (*i.e.* new classification results), can be extracted to combine with the clustering in the next iteration. Reliable information from each individual classification can be extracted as the iterative process converges.

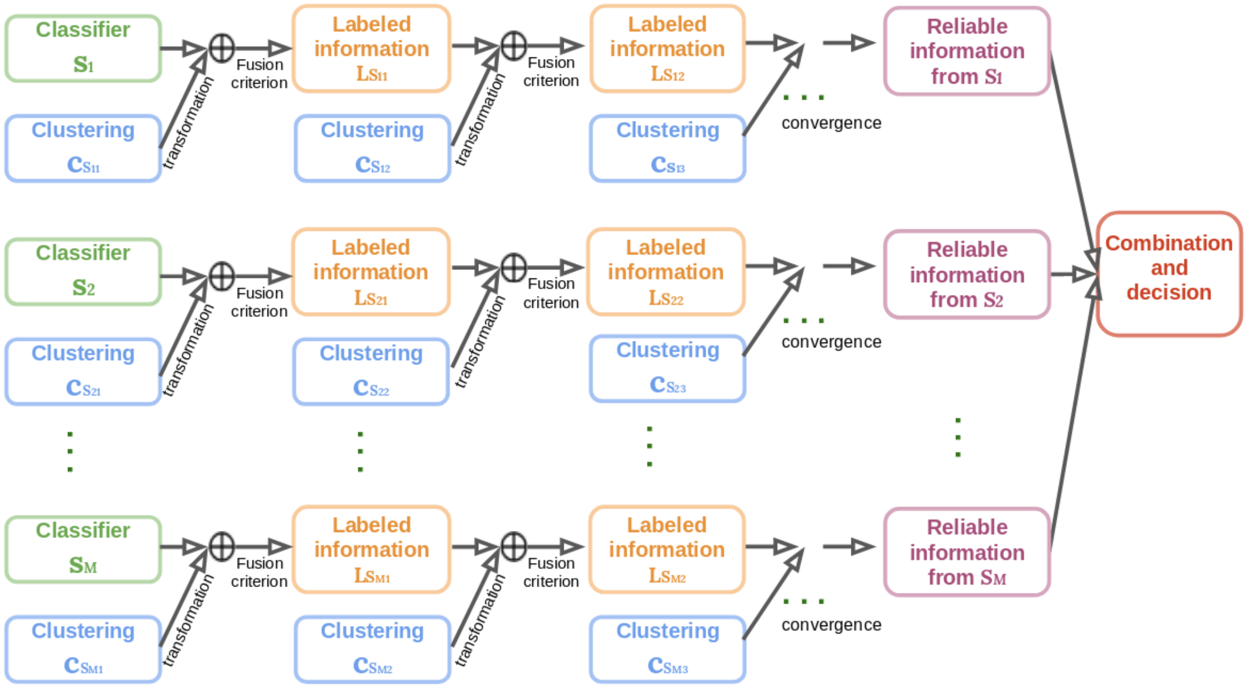


Figure 2: Workflow of the proposed EFSC.

In the framework of belief functions, uncertainty and imprecision are separately represented by BBA values on singletons and on unions. After the transformation, information from clustering can be thus considered as a supplementary source to add information to classification. For a classification, the combination of clustering by Dempster's rule helps to reinforce the support on the same focal element, which thus increase the corresponding BBA value. A higher BBA on certain focal element indicates more belief degrees are assigned and the decision on this focal element thus become more certain. Furthermore, a combined BBA after Dempster's combination rule has less uncertainty or imprecision, so that it further approaches to the categorical BBA which represents the perfect information with neither uncertainty nor imprecision.

We thus use Jousselme distance between a BBA and the corresponding categorical BBA to represent the loss of confidence. That is to say, the combination reduces uncertianty and imrpecsion, so that makes the BBA more approaching to the perfect information, leading to a smaller Jousselme distance. Therefore, the iterative combination with multiple clustering results by Dempster's rule can gradually reduce uncertainty and imprecision in the former BBA. The criterion based on Jousselme distance to select partial information can control the loss of confidence after combining with clustering.

6. Numerical example

In this section, we show a numerical example to explain the proposed transformation and the iterative fusion process. For a group of objects with eight elements, noted as $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, their corresponding ground truth, classification and clustering results are detailed in Table 1. The classification result has the same frame of discernment as the ground truth, *i.e.* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. We have five different clustering results on X, separating objects into either four clusters such as c_1 or three clusters such as c_2, c_3, c_4 and c_5. To combine the classification and clustering results, we have first to transfer the BBAs of clustering methods to the same frame as classification.

Table 1: Labels of ground truth, classification and clustering results on objects in X.

Labels	Objects in X								Frame of discernment
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
Ground truth	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
s_1	ω_4	ω_1	ω_2	ω_1	ω_2	ω_3	ω_2	ω_4	$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
c_1	θ_{11}	θ_{11}	θ_{12}	θ_{12}	θ_{12}	θ_{13}	θ_{14}	θ_{14}	$\Theta_1 = \{\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}\}$
c_2	θ_{23}	θ_{21}	θ_{21}	θ_{21}	θ_{21}	θ_{22}	θ_{23}	θ_{23}	$\Theta_2 = \{\theta_{21}, \theta_{22}, \theta_{23}\}$
c_3	θ_{33}	θ_{33}	θ_{31}	θ_{31}	θ_{31}	θ_{32}	θ_{33}	θ_{33}	$\Theta_3 = \{\theta_{31}, \theta_{32}, \theta_{33}\}$
c_4	θ_{41}	θ_{41}	θ_{42}	θ_{42}	θ_{43}	θ_{43}	θ_{42}	θ_{42}	$\Theta_4 = \{\theta_{41}, \theta_{42}, \theta_{43}\}$
c_5	θ_{51}	θ_{51}	θ_{52}	θ_{52}	θ_{52}	θ_{53}	θ_{53}	θ_{53}	$\Theta_5 = \{\theta_{51}, \theta_{52}, \theta_{53}\}$

6.1. Construction of BBAs of classification and clustering results

This numerical example simulates a case where raw data of objects is not available, and we only have the results of classification and clustering. The BBAs of classification are therefore generated randomly on 2^Ω without accessing to the raw data, as shown in Table 2. As all possible states of objects labels are included in Ω, the BBAs of classification are constructed under the closed-world assumption satisfying $m(\emptyset) = 0$. The classification results should indicate that decisions are on the singletons corresponding to the current labels. Therefore, all subsets except the empty set in Ω are assigned with random value by the uniform distribution, and the singletons corresponding to the current label have the maximum BBAs, marked in bold in Table 2.

The clustering results are on the frames Θ_1, Θ_2, Θ_3, Θ_4 and Θ_5 and their BBAs are successively noted as $m_{c_1}^{\Theta_1}$ for clustering c_1, $m_{c_2}^{\Theta_2}$ for clustering c_2, $m_{c_3}^{\Theta_3}$ for clustering c_3, $m_{c_4}^{\Theta_4}$ for clustering c_4, $m_{c_5}^{\Theta_5}$ for clustering c_5. We use the model presented in section 3.4 to estimate the BBA of a clustering method by a group of SBBAs. For each object $x_i \in X$, to simplify the calculation, its SBBA related to the cluster to which x_i belongs, is constructed with 0.8 on the singleton and 0.2 on the ignorance. For other clusters that x_i does not belong to, the corresponding SBBAs have 1 on the ignorance, representing we knowing nothing about them. Therefore, the BBA of x_i can be presented by only one SBBA with the cluster it belonging to as the focal element. For example, for the object x_1 belonging to the cluster θ_{11} in c_1, its SBBA is defined as:

$$\begin{cases} m_{c_1}^{\Theta_1}(\theta_{11})(x_1) = 0.8, \\ m_{c_1}^{\Theta_1}(\Theta_1)(x_1) = 0.2. \end{cases} \tag{32}$$

Table 2: BBAs of classification

2^Ω	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\emptyset$	0.	0.	0.	0.	0.	0.	0.	0.
ω_1	0.0578	**0.1271**	0.0616	**0.1212**	0.1274	0.0864	0.0781	0.0183
ω_2	0.0187	0.0737	**0.0941**	0.0258	**0.1679**	0.0135	**0.1291**	0.0221
$\omega_1 \cup \omega_2$	0.0255	0.0650	0.0558	0.1045	0.0336	0.0598	0.0611	0.0591
ω_3	0.0642	0.0502	0.0561	0.1031	0.0650	**0.1164**	0.0700	0.0939
$\omega_1 \cup \omega_3$	0.0588	0.0739	0.0622	0.1198	0.0272	0.0842	0.1019	0.0272
$\omega_2 \cup \omega_3$	0.0702	0.0606	0.0165	0.0792	0.1023	0.0648	0.1258	0.0366
$\omega_1 \cup \omega_2 \cup \omega_3$	0.0936	0.0250	0.0622	0.0138	0.0725	0.0376	0.1001	0.0906
ω_4	**0.1122**	0.0060	0.0501	0.0485	0.0377	0.0078	0.0281	**0.1149**
$\omega_1 \cup \omega_4$	0.1098	0.0072	0.0886	0.0258	0.0207	0.0993	0.0157	0.0515
$\omega_2 \cup \omega_4$	0.0915	0.0630	0.0901	0.1056	0.0468	0.0767	0.0433	0.1017
$\omega_1 \cup \omega_2 \cup \omega_4$	0.1086	0.0626	0.0920	0.0253	0.1212	0.0587	0.0618	0.0322
$\omega_3 \cup \omega_4$	0.0401	0.0865	0.0344	0.0445	0.0139	0.1107	0.0915	0.1085
$\omega_1 \cup \omega_3 \cup \omega_4$	0.0238	0.1065	0.0835	0.0713	0.0496	0.0647	0.0466	0.0900
$\omega_2 \cup \omega_3 \cup \omega_4$	0.0899	0.1042	0.0623	0.0289	0.0321	0.1064	0.0791	0.0936
Ω	0.0345	0.0877	0.0896	0.0819	0.0815	0.0122	0.0389	0.0589

6.2. Transformation of BBAs

In this section, we explain how to transfer the BBAs of clustering c_1 defined on Θ_1 to the frame of discernment Ω. We measure the similarity between the results of classification s_1 and clustering c_1 by

Jaccard index, as shown in Table 3.

Table 3: Similarity of classes and clusters measured by Jaccard index.

Classes	Clusters from clustering c_1			
	θ_{11}	θ_{12}	θ_{13}	θ_{14}
ω_1	0.333	0.25	0	0
ω_2	0	0.5	0	0.25
ω_3	0	0	1	0
ω_4	0.333	0	0	0.333

Let's take the object x_1 again as an example. The clustering label of x_1 is cluster θ_{11} which has intersections with two classes: ω_1 and ω_4. Thus we take the similarities between cluster θ_{11} and class ω_1, and also class ω_4 into account to achieve the transformation of $m_{c_1}^{\Theta_1}(x_1)$. The value of similarity 0.333 between cluster θ_{11} and class ω_1 indicates the belief degree that class ω_1 and cluster θ_{11} are the same, and nothing more. As illustrated in equation (17), the similarity can be represented by a BBA on the frame of discernment H, defined as:

$$\begin{cases} m_{\omega_1,\theta_{11}}^{H}(Y) = 0.333, \\ m_{\omega_1,\theta_{11}}^{H}(N) = 1 - 0.333 = 0.667. \end{cases} \tag{33}$$

The transformation is composed of four major steps as explained in section 4.2. The first step is to decompose the BBA of clustering into a group of SBBAs considered as multiple independent sources. For object x_1, its BBA $m_{c_1}^{\Theta_1}(x_1)$ are already in the form of SBBA with $\{\theta_{11}\}$ as the focal element, as shown in equation (32), and thus can be directly used. The second step is to transfer the SBBAs to Ω by the corresponding similarity between cluster θ_{11} and all possible classes with non-empty intersections, *i.e.* ω_1 and ω_4. According to equation (20) and equation (21), the transformed SBBA of x_1 on ω_1 is calculated as:

$$\begin{aligned} m_{\theta_{11},\omega_1}^{\Omega}(\omega_1)(x_1) &= m_{\omega_1,\theta_{11}}^{H}(Y) * m_{c_1}^{\Theta_1}(\theta_{11})(x_1) \\ &= 0.333 * 0.8 = 0.2664 \end{aligned} \tag{34}$$

$$\begin{aligned} m_{\theta_{11},\omega_1}^{\Omega}(\Omega)(x_1) &= 1 - m_{\omega_1,\theta_{11}}^{H}(Y) * m_{c_1}^{\Theta_1}(\theta_{11})(x_1) \\ &= 0.7336 \end{aligned} \tag{35}$$

The transformed SBBA of x_1 on ω_4 is calculated as:

$$\begin{aligned} m_{\theta_{11},\omega_4}^{\Omega}(\omega_4)(x_1) &= m_{\omega_4,\theta_{11}}^{H}(Y) * m_{c_1}^{\Theta_1}(\theta_{11})(x_1) \\ &= 0.333 * 0.8 = 0.2664 \end{aligned} \tag{36}$$

$$\begin{aligned} m_{\theta_{11},\omega_4}^{\Omega}(\Omega)(x_1) &= 1 - m_{\omega_4,\theta_{11}}^{H}(Y) * m_{c_1}^{\Theta_1}(\theta_{11})(x_1) \\ &= 0.7336 \end{aligned} \tag{37}$$

Table 4: Transformed BBAs from clustering c_1 on Ω.

2^Ω	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\emptyset$	0.	0.	0.	0.	0.	0.	0.	0.
ω_1	0.2105	0.2105	0.1304	0.1304	0.1304	0.	0.	0.
ω_2	0.	0.	0.3478	0.3478	0.3478	0.	0.1549	0.1549
$\omega_1 \cup \omega_2$	0.	0.	0.	0.	0.	0.	0.	0.
ω_3	0.	0.	0.	0.	0.	0. 8	0.	0.
$\omega_1 \cup \omega_3$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_2 \cup \omega_3$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_1 \cup \omega_2 \cup \omega_3$	0.	0.	0.	0.	0.	0.	0.	0.
ω_4	0.2105	0.2105	0.	0.	0.	0.	0.2253	0.2253
$\omega_1 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_2 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_1 \cup \omega_2 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_3 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_1 \cup \omega_3 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
$\omega_2 \cup \omega_3 \cup \omega_4$	0.	0.	0.	0.	0.	0.	0.	0.
Ω	0.5789	0.5789	0.5217	0.5217	0.5217	0.2	0.6197	0.6197

The third step is to combine all available SBBAs on each classes, as illustrated in equation (27). As we only have one SBBA$m^\Omega_{\theta_{11},\omega_1}(x_1)$ on ω_1 and one SBBA $m^\Omega_{\theta_{11},\omega_4}(x_1)$ on ω_4, thus we have:

$$m^\Omega_{c_1,\omega_1}(x_1) = m^\Omega_{\theta_{11},\omega_1}(x_1) \tag{38}$$

$$m^\Omega_{c_1,\omega_4}(x_1) = m^\Omega_{\theta_{11},\omega_4}(x_1) \tag{39}$$

where $m^\Omega_{c_1,\omega_1}(x_1)$ represents that the information of object x_1 from clustering c_1 is transformed on ω_1 and $m^\Omega_{c_1,\omega_4}(x_1)$ is transformed on ω_4. The last step is to combine all possible SBBAs on Ω transformed from clustering c_1 by equation (28) as:

$$\begin{aligned} m^\Omega_{c_1}(x_1) =& m^\Omega_{c_1,\omega_1}(x_1) \oplus m^\Omega_{c_1,\omega_4}(x_1) \\ &\Rightarrow \begin{cases} m^\Omega_{c_1}(\omega_1)(x_1) = 0.2105, \\ m^\Omega_{c_1}(\omega_4)(x_1) = 0.2105, \\ m^\Omega_{c_1}(\Omega)(x_1) = 0.5789, \\ m^\Omega_{c_1}(A)(x_1) = 0, \forall A \subset \Omega, A \neq \omega_1, A \neq \omega_4. \end{cases} \end{aligned} \tag{40}$$

For other objects, their transformed BBAs on Ω are shown in Table 4.

6.3. Iterative fusion process

After the transformation, the information from clustering c_1 can be combined with the BBAs of classification presented in Table 2, and the combined results of all objects are shown in Table 5. Decisions are made on the combined BBAs by equation (13), and the loss of confidence of each decision is represented by the method from [12] based on Jousselme distance from its BBAs to the corresponding categorical BBAs by equation (14), as shown in Table 6.

Table 5: BBAs on Ω after combination of clustering c_1 and classification s_1.

2^Ω	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\emptyset$	0.	0.	0.	0.	0.	0.	0.	0.
ω_1	0.1735	0.2394	0.1385	0.1824	0.1696	0.0257	0.0624	0.0117
ω_2	0.0132	0.0532	0.3100	0.2295	0.3979	0.0028	0.2353	0.1051
$\omega_1 \cup \omega_2$	0.0181	0.0466	0.0359	0.0731	0.0186	0.0174	0.0473	0.0431
ω_3	0.0456	0.0353	0.0361	0.0721	0.0409	0.7418	0.0552	0.0699
$\omega_1 \cup \omega_3$	0.0417	0.0534	0.0404	0.0840	0.0140	0.0250	0.0833	0.0185
$\omega_2 \cup \omega_3$	0.0498	0.0432	0.0080	0.0550	0.0674	0.0189	0.1044	0.0257
$\omega_1 \cup \omega_2 \cup \omega_3$	0.0664	0.0160	0.0404	0.0081	0.0463	0.0104	0.0817	0.0674
ω_4	0.2373	0.1382	0.0319	0.0329	0.0215	0.0010	0.1063	0.2617
$\omega_1 \cup \omega_4$	0.0780	0.0024	0.0592	0.0168	0.0094	0.0298	0.0074	0.0372
$\omega_2 \cup \omega_4$	0.0650	0.0451	0.0603	0.0738	0.0280	0.0227	0.0317	0.0759
$\omega_1 \cup \omega_2 \cup \omega_4$	0.0771	0.0447	0.0617	0.0164	0.0808	0.0170	0.0479	0.0224
$\omega_3 \cup \omega_4$	0.0284	0.0630	0.02076	0.0301	0.0046	0.0334	0.0107	0.0811
$\omega_1 \cup \omega_3 \cup \omega_4$	0.0169	0.0783	0.0556	0.0493	0.0300	0.0189	0.0346	0.0669
$\omega_2 \cup \omega_3 \cup \omega_4$	0.0638	0.0765	0.0405	0.0189	0.0176	0.0321	0.0632	0.0697
Ω	0.0245	0.0639	0.0600	0.0569	0.0527	0.0024	0.0278	0.0430

The labels after decision should be compared with the original classification labels according to the proposed criterion in algorithm 1, to decide whether the information from clustering is finally added or not. For each class, we update their labels if the average loss of confidence of this class after the combination is reduced, compared to its counterpart calculated from the BBAs of classification s_1, also shown in Table 6.

Let's take the class ω_1 as an example. Its average loss of confidence in s_1, noted as $\overline{R_{\omega_1,s_1}}$, is calculated as:

$$\begin{aligned}\overline{R_{\omega_1,s_1}} &= \frac{R_{x_2,s_1}+R_{x_4,s_1}}{2}\\ &= \frac{0.638+0.612}{2}\\ &= 0.625\end{aligned} \tag{41}$$

where R_{x_2,s_1} and R_{x_4,s_1} represent the loss of confidence of objects x_2 and x_4 in s_1. After the combination with clustering c_1, the decisions on the combined BBAs update the label and also the loss of confidence of each object. Therefore, for the class ω_1, its new average loss of confidence $\overline{R_{\omega_1}}$ is calculated as:

$$\overline{R_{\omega_1}} = R_{x_2} = 0.565 \tag{42}$$

where R_{x_2} is the loss of confidence of object x_2 after the combination. As we have $\overline{R_{\omega_1}} < \overline{R_{\omega_1,s_1}}$, the information on ω_1 from the original classification s_1 is thus updated by the new one after combination. That is to say, the class ω_1 in s_1 originally contains two objects x_2 and x_4 is updated as only one object x_2 according to the results after combination.

Table 6: Decisions on combined BBAs (compared with the labels and loss of confidence of classification s_1).

Decisions	Objects in X							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Labels	ω_4	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4
Losses of confidence	0.543	0.565	0.508	0.553	0.428	0.178	0.528	0.514
Labels of s_1	ω_4	ω_1	ω_2	ω_1	ω_2	ω_3	ω_2	ω_4
Losses of confidence of s_1	0.613	0.638	0.650	0.612	0.573	0.606	0.5969	0.605

The newly updated labels can be considered as a new classification result to combine with another randomly selected clustering result among c_1, c_2, c_3, c_4 and c_5. In Table 7, we show the details of the proposed iterative fusion process, including the clustering used for combination, the average loss of confidence on each class (*i.e.* $\overline{R_{\omega_1}}$, $\overline{R_{\omega_2}}$, $\overline{R_{\omega_3}}$, $\overline{R_{\omega_4}}$), the global average loss of confidence $\overline{R}$, and the labels of objects in each iteration. We run ten iterations in total and the average loss of confidence $\overline{R}$ gradually decreases from 0.611 in classification s_1 to 0.002 after the combination with multiple clustering results. The original accuracy of s_1 is 0.625, and it gradually increases to 1 after the fifth iteration.

Table 7: Iterative fusion process

Combination	Losses of confidence					Labels								Accuracy
	$\overline{R_{\omega_1}}$	$\overline{R_{\omega_2}}$	$\overline{R_{\omega_3}}$	$\overline{R_{\omega_4}}$	$\overline{R}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
s_1	0.625	0.606	0.606	0.608	0.611	ω_4	ω_1	ω_2	ω_1	ω_2	ω_3	ω_2	ω_4	0.625
$+c_1$	0.565	0.504	0.178	0.527	0.443	ω_4	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.738
$+c_1$	0.496	0.320	0.039	0.441	0.324	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.863
$+c_2$	0.485	0.255	0.008	0.451	0.299	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.863
$+c_3$	0.421	0.194	0.002	0.436	0.263	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_2	ω_4	0.863
$+c_1$	0.126	0.031	0.000	0.425	0.145	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_1$	0.028	0.006	0.000	0.138	0.057	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_4$	0.005	0.005	0.000	0.114	0.041	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_1$	0.001	0.001	0.000	0.025	0.006	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_5$	0.000	0.000	0.000	0.013	0.003	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0
$+c_1$	0.000	0.000	0.000	0.009	0.002	ω_1	ω_1	ω_2	ω_2	ω_2	ω_3	ω_4	ω_4	1.0

7. Experiments

In this section, we evaluate the proposed EFSC on synthetic data and on real remote sensing data from different aspects. With synthetic data, we principally discuss the performance of EFSC from the aspect at the output level, by controlling the quality of its direct inputs: classification and clustering results. With the real data, the evaluation of EFSC focuses more on the information at the raw data level (*e.g.* the quality of training samples of classifiers). To make the experiments concise and clear, we configure EFSC into three different ways:

1. **EFSC11**: one classification and one clustering methods.
2. **EFSC1m**: one classification and multiple clustering methods.
3. **EFSCmm**: Multiple classification and multiple clustering methods.

We discuss the performance of EFSC on synthetic data in section 7.1. EFSC focuses on the fusion of classification or clustering results rather than raw data, features nor supervised/unsupervised classifiers themselves. This indicates that the direct inputs of EFSC are the classification and clustering results and their corresponding BBAs. Therefore, we verify EFSC with the synthetic data at the output level in a controlled environment. In section 7.1.2, we study how the quality of classification and clustering results affect the combination for the EFSC11 configuration. In section 7.1.3, we focus on the EFSC1m configuration and study the same questions as section 7.1.2. In section 7.1.4, we evaluate EFSCmm configuration with multiple classification and clustering methods and compare the results with other fusion approaches.

On the real remote sensing datasets, we conduct four experiments to study: (1) how clustering heterogeneity affects the combination; (2) whether multiple clustering methods can reduce the uncertainty of a supervised result; (3) how mislabeled training samples affect the fusion results; (4) how the EFSC behaves during fusion with multiple supervised and clustering methods.

In the first experiment in section 7.2.2, we test the EFSC11 configuration and change the numbers of partitions in clustering. In section 7.2.3, the second experiment studies the proposed iterative fusion process. We focus on the EFSC1m configuration to demonstrate that information from multiple clustering methods can progressively reduce the uncertainty and increase the accuracy of the supervised results until convergence. In section 7.2.4, we further investigate the robustness of the EFSC1m configuration for the mislabeled training samples in an artificial case and a real situation. The last experiment in section 7.2.5 focuses on the efficiency of the EFSCmm configuration where multiple supervised results were combined individually with multiple clustering methods. The extracted reliable information then are combined to improve the overall accuracy. EFSCmm is constructed based on EFSC1m whose effectiveness in reducing uncertainty and robustness for mislabeled training samples has already been discussed in sections 7.2.3 and 7.2.4. For EFSCmm, we therefore focus only on its performance to improve the overall accuracy.

We select two distinct fusion methods at the production level, based on different principles. EC3 belongs to the group of fusion methods that optimize the agreements between supervised and pooled results, and it was shown to be more efficient than the other methods in this category [6]. Similar to our work, the method proposed by Karem allows to obtain a combination through belief functions.

7.1. Experiments on synthetic data

7.1.1. Synthetic data at the output level

The synthetic data at the output level includes the outputs of supervised or unsupervised classifiers and their corresponding BBAs, which are used as inputs of EFSC. To study how qualities of classification and clustering results affect the combination, we generate an image as ground truth, and gradually add uniformly distributed mistakes on it to obtain classification and clustering results. In this way, we can directly control the quality of supervised and unsupervised results. The experiments on synthetic labels simulate a case where raw data is inaccessible, and we only have the results of classification and clustering. The experiments on synthetic data are all launched 15 times to obtain the averages.

Table 8: Descriptions of the synthetic area.

Labels	Forest	Shrub/Scrub	Grassland /Herbaceous	Wetlands	Total
Synthetic area	20687	11417	3526	4370	40000

The process to generate an classification result with the proportion of mistakes M_s based on the ground truth is detailed as:

1) Randomly select pixels with proportion M_s per class;
2) For the selected pixels, change their labels randomly into any other class.

The random BBAs of the synthetic classification result on the frame of discernment Ω are constructed as:

1) For a pixel x labeled as ω_i, we have $m(\omega_i) = 1 - M_s$;
2) The rest of BBAs are randomly set to $[0, 1 - M_s)$;
3) Normalize the BBAs to satisfy equation (1).

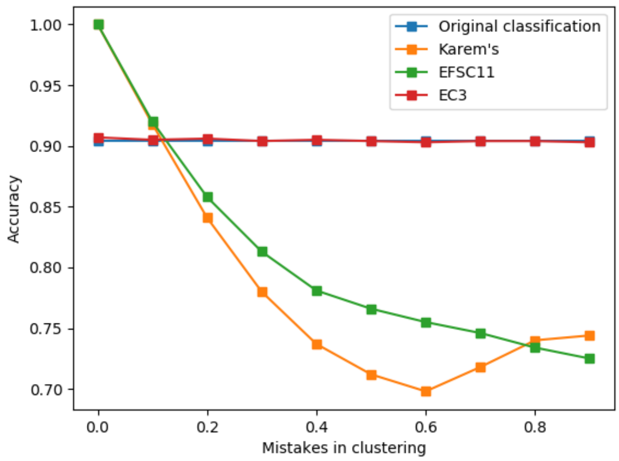
Generating a synthetic clustering result includes two major steps: separating or gathering classes into clusters, and add noise with proportion M_c into each cluster to reduce its homogeneity. The quality of clustering indicates whether data are well-separated, which is different from the evaluation of classification. We use the homogeneity in clusters to represent the quality of clustering results, which thus can be controlled by M_c. For example, if a class on the ground truth is separated into two clusters, the homogeneity in each cluster is not reduced, and thus we still can consider it as a perfect clustering. After adding noise with M_c, the homogeneity in each cluster is gradually reduced. We note the number of classes on the ground truth as n, and the number of clusters as k. The process to generate a clustering result is as:

1) If $k = n$, classes on the ground truth is directly take as clusters;
2) If $k < n$, we randomly select $n - k$ classes on the ground truth and gather them as one cluster. The rest classes are kept as clusters;
3) If $k > n$, we use an iterative process to add clusters. In each iteration, we randomly select a cluster and separate into two clusters until the clusters are enough. The initial state of this process is the ground truth whose classes are directly take as clusters;
4) For the clustering result generated previously, we randomly select pixels in each cluster with the proportion M_c and change its labels to any other cluster.

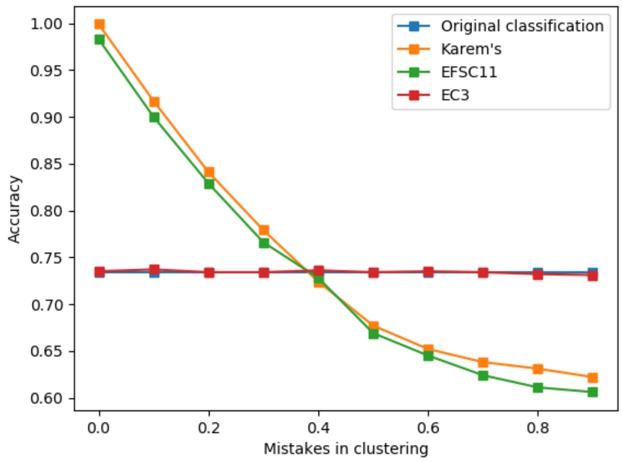
The BBAs of clustering on Θ are constructed in the form of SBBAs. For a pixel x in the cluster θ_j, we have $m(\theta_j) = 1 - M_c$ and $m(\Theta) = M_c$.

7.1.2. Combination of one classification and one clustering methods

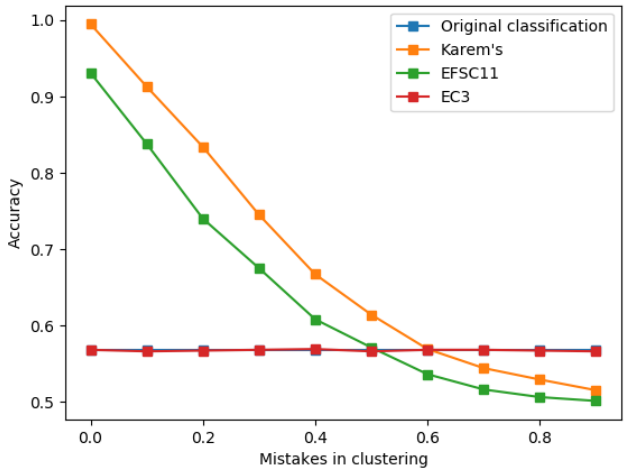
In this section, we evaluate EFSC11 configuration by combining one classification and one clustering, compared with Karem's and EC3. We focus on how quality of classification or clustering affect the combination. For the clustering, we fix its number of clusters $k = 7$ and gradually add mistakes with proportion $M_c \in [0, 1]$. As we use the homogeneity to represent the quality of a clustering, $M_c = 0$ represents the perfect case where no mistakes are involved. With M_c increasing, the homogeneity of each cluster begins to decrease, which thus can be considered as the decline of the clustering quality.



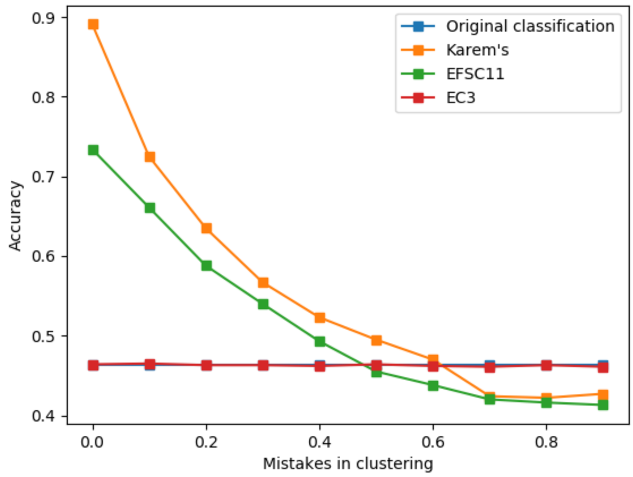
(a) $M_s = 0.1$ (accuracy of classification = 0.904).



(b) $M_s = 0.3$ (accuracy of classification = 0.734).



(c) $M_s = 0.5$ (accuracy of classification = 0.586).



(d) $M_s = 0.7$ (accuracy of classification = 0.463).

Figure 3: Accuracy change with mistakes in clustering for the combination of one classification and one clustering.

We fix the mistakes M_s in classification as 0.1, 0.3, 0.5 and 0.7 and for each of them, M_c varies from 0 to 0.9, as shown in Figure 3. A perfect clustering with $M_c = 0$ increases the accuracy of classification, whereas the accuracy after combinations by EFSC11 and Karem's are not improved when $M_c \geq 0.2$, as shown in Figure 3a. The BBAs of classification are assigned by random values but also guarantee the singletons corresponding to the classification labels have the maximum BBA. Therefore, for the classification with $M_s = 0.1$, even though almost all pixels have true labels, the BBAs are still constructed with uncertainty and imprecision, indicating that the classification is good enough yet not completely reliable. Clustering information can be thus partially taken into account, however possibly leading to decreasing the accuracy when the quality of clustering is relatively poorer than classification. We can also roughly observe this in Figure 3b, 3c and 3d. When $M_s = 0.3$, the accuracy after combination is improved only if $M_c < 0.4$, and for $M_s = 0.5$ and $M_s = 0.7$, improvement occurs when $M_c < 0.5$.

We also fix M_c for clustering and change the mistakes M_s in classification. To avoid the influence from mistakes in clustering, we select a good enough clustering with $M_c = 0.1$, and study how the quality of classification affects the combination. The results are shown in Figure 4. We can observe that combining with a good enough clustering can noticeably enhance the accuracy, even for the low quality classification (*e.g.*, $M_s = 0.7$). When $M_s > 0.8$, it indicates that the classification is almost incorrect and thus the combination is helpless to improve the accuracy, even with a good enough clustering.

In the combination of one classification and one clustering, the enhancement of accuracy is still pronounced after combination by EFSC11 and Karem's. Nevertheless, Karem's has better performance than EFSC11 because EFSC11 takes information more prudently so that less information is fused compared to Karem's. Accordingly, Karem's is more pertinent for combining one classification and one clustering, whereas EFSC has more advantage when multiple classification and clustering are available. EC3 has the similar performance as the original classification because it relies more on classification than clustering. EC3 is more pertinent for multiple classification results with high qualities.

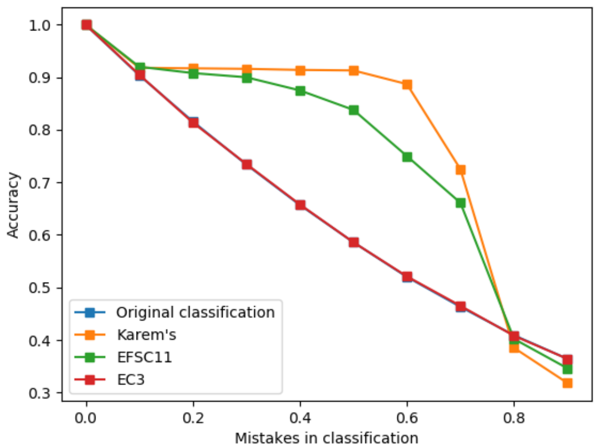
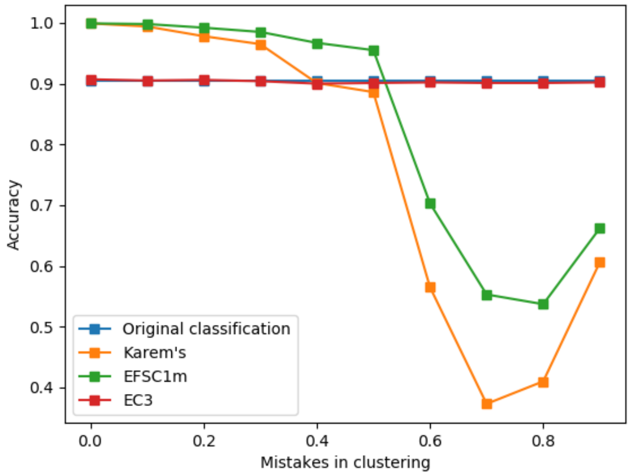


Figure 4: Accuracy change with mistakes in classification for the combination of one classification and one clustering.

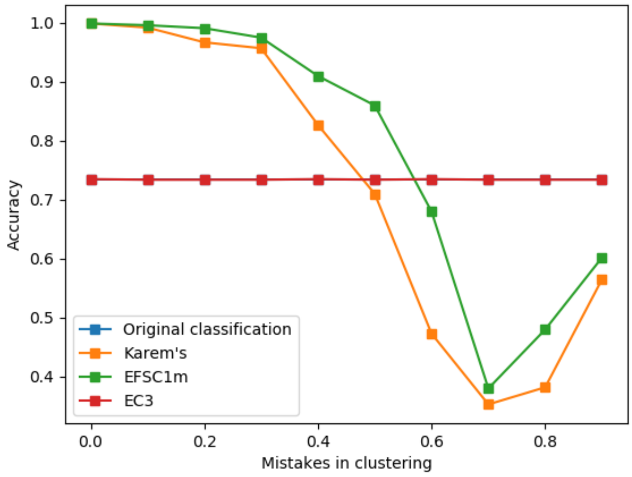
7.1.3. Combination of one classification and multiple clustering methods

In this section, we evaluate the EFSC1m configuration with one classification and multiple clustering methods. We also fix M_s separately as 0.1, 0.3, 0.5 and 0.7 and gradually change M_c from 0 to 0.9. A group of clustering methods has the number of clusters varying from 3 to 15.

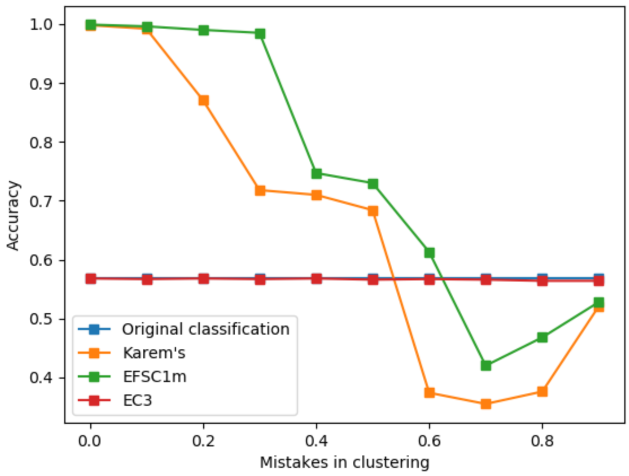
For EFSC1m and Karem's methods, combination with multiple clustering methods can evidently improve the accuracy of classification, as shown in Figure 5. A high-quality classification boosts the combination results with multiple clustering, such as results in Figures 5a, 5b.



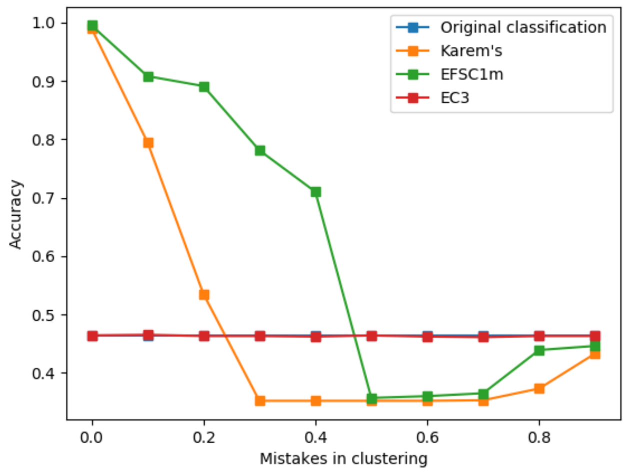
(a) $M_s = 0.1$ (accuracy of classification $= 0.904$).



(b) $M_s = 0.3$ (accuracy of classification $= 0.734$).



(c) $M_s = 0.5$ (accuracy of classification $= 0.586$).



(d) $M_s = 0.7$ (accuracy of classification $= 0.463$).

Figure 5: Accuracy change with mistakes in clustering for the combination of one classification and multiple clustering methods.

Even if the mistakes in classification are considerable, *i.e.* $M_s = 0.7$, EFSC1m and Karem's methods can still increase the overall accuracy, as shown in Figure 5d. We can observe that combination with multiple clustering methods can better decrease the influence by the quality of classification, compared with the results in the previous section. When M_s reaches 0.1, 0.3 and 0.5, the accuracy of classification is improved when $M_c < 0.5$. This indicates that when combining with multiple clustering methods, EFSC1m and Karem's methods are easily influenced by clustering than classification.

When $M_c > 0.6$, the accuracy after combination slightly increases yet always worse than the original classification. This increase occurs when clustering evidently has low quality because the BBAs of clustering, in this case, are mainly distributed on the total ignorance. Therefore, information from clustering with low quality is rarely combined with classification, consequently avoiding further reduction of accuracy.

We also fix the quality of clustering with $M_c = 0.1$ and study the accuracy change with M_s, as shown in Figure 6. The accuracy after combination shows an evident decline when $M_c > 0.5$. This result also proved that when the quality of clustering is acceptable, the combination is less influenced by classification.

In the combination of one classification and multiple clustering methods, EFSC1m outperforms the other two methods because it cautiously takes into account the information from clustering. EC3 still shows similar performance as the original classification because when maximizing the consensus of classification and clustering, only one classification is available to provide the semantic labels. This indicates that EC3 relies more on classification than clustering. Compared to EC3, EFSC1m and Karem's methods with the proposed iterative fusion process can well handle the combination with one classification and multiple clustering methods.

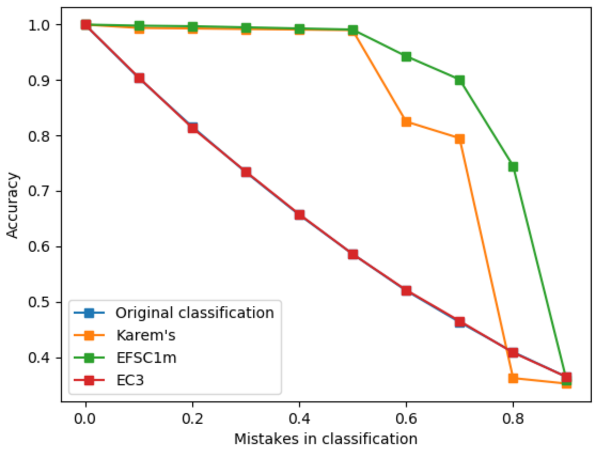


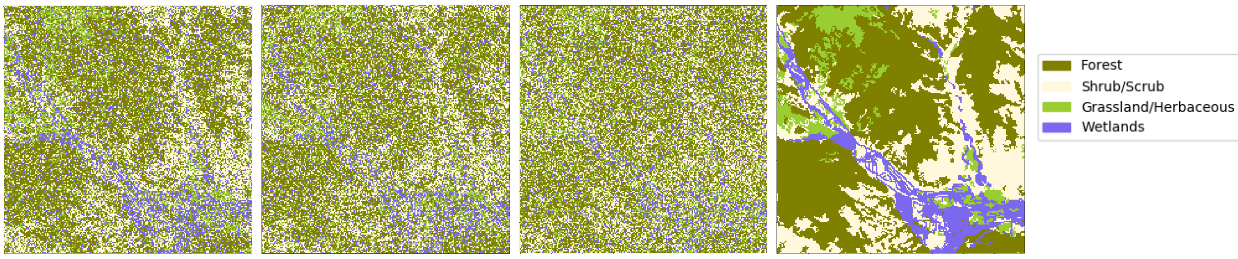
Figure 6: Accuracy change with mistakes in classification for the combination of one classification and multiple clustering methods.

7.1.4. Combination of multiple classification and clustering methods

In this section, we combine multiple classification and clustering results separately by EFSCmm, Karem's and EC3 methods, also compared with the fusion of multiple classifications by Majority Voting (MV). We randomly generate three classification results with the accuracy of 0.586, 0.522, 0.463. The group of clustering methods has the number of partitions from 3 to 15 with $M_c = 0.3$.

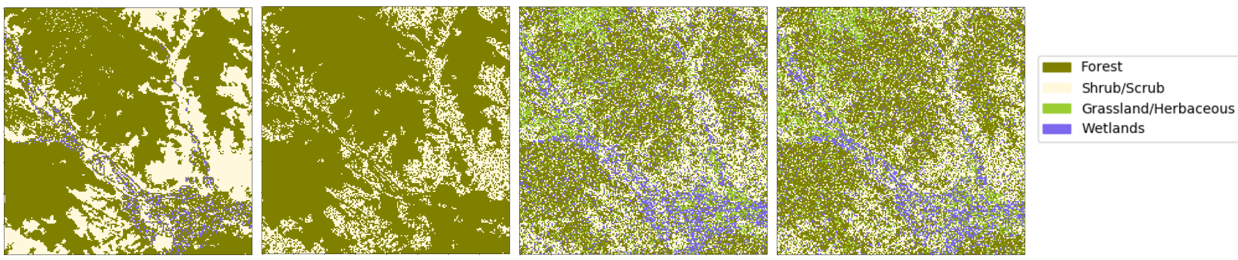
Table 9 shows the overall accuracy measured by F1 score and also for each class. We also show the classification and combination results in Figure 8. EFSCmm can noticeably enhance the accuracy by 30% at most compared to the best classification, and it also surpasses other combination methods. Karem's also shows an evident improvement of the overall accuracy, whereas totally ignores two classes: *Grassland/Herbaceous* and *Wetlands*. Although EFSCmm can not well classify *Grassland/Herbaceous* either, it improves the accuracy of *Wetlands*, demonstrating it is more prudent than Karem's. EC3 performs better with three classifications compared to its previous performance, because it depends more on classification than clustering. For the combination results of the three classifications by MV, we can observe it is less pertinent than the other three combination methods, indicating the importance of clustering information.

We assume that for each object x, the probability that an object is correctly classified follows Gaussian distribution. Thus for 40000 pixels, each of them can be regard as an independent result. We thus consider approximately that each test is run 40000×15 to calculate the confidence interval. The confidence interval with the confidence level as 95% of EFSCmm is $[0.887, 0.890]$, Karem's $[0.793, 0.796]$, EC3 $[0.602, 0.605]$, and MV $[0.593, 0.596]$. The experiments show that EFSCmm is significantly better than Karem's, EC3 and MV on synthetic data.



(a) Classification 1 (b) Classification 2 (c) Classification 3 (d) Ground truth

Figure 7: Synthetic classification results and ground truth.



(a) EFSCmm (b) Karem's (c) EC3 (d) MV

Figure 8: Combination results of EFSCmm, Karem's, EC3 and MV on synthetic data.

Table 9: Accuracy of three original classifications and four fusion methods (Karem's, EFSCmm, EC3, MV) on synthetic data.

Methods	Accuracy per class				Accuracy
	Forest	Shrub /Scrub	Grassland /Herbaceous	Wetlands	
Classification 1	0.667	0.573	0.363	0.421	0.586
Classification 2	0.609	0.497	0.292	0.355	0.522
Classification 3	0.562	0.422	0.230	0.285	0.463
Karem's	0.786	0.828	0.	0.	0.795
EFSCmm	**0.880**	**0.955**	0.026	0.433	**0.889**
EC3	0.682	0.590	**0.389**	0.446	0.604
MV	0.705	0.594	0.388	**0.452**	0.595

7.2. Experiments on real remote sensing data sets

7.2.1. Data description and study areas

The study area is located in Colorado, USA, and contains two national forest parks with a variety of vegetation, as shown in Figure 3a. The satellite data used in our experiment are from LandSat-8 OLI. It consists of eight spectral bands with a spatial resolution of 30 meters, a panchromatic band with a resolution of 15 meters and two thermal bands with a resolution of 100 meters. The remote sensing data acquired 11^{th} June 2018 were obtained from the USGS Earth Explorer. The geometric correction of the image was performed by the transformation of the UTM map by NASA. We used the multispectral bands $1-7$ in our test. For ground truth, we use the labels generated by National Land Cover Database (NLCD) 2016 and group its land cover pattern at the basic level: *Water, Developed Area, Forest, Shrub/Scrub, Grassland/Herbaceous, Pasture/Crops, Wetlands.* We randomly select two test areas, one with 2500 pixels and the other with 40000 pixels, detailed in Figure 9 and Table 10. Most of the experiments are conducted on the test area 1, but we also evaluate the proposed EFSC on the test area 2 in section 7.2.5.

Table 10: Descriptions of the test areas.

	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	Total
Test area 1	146	182	266	1558	15	148	185	2500
Test area 2	77	571	21147	13609	593	2250	1753	40000

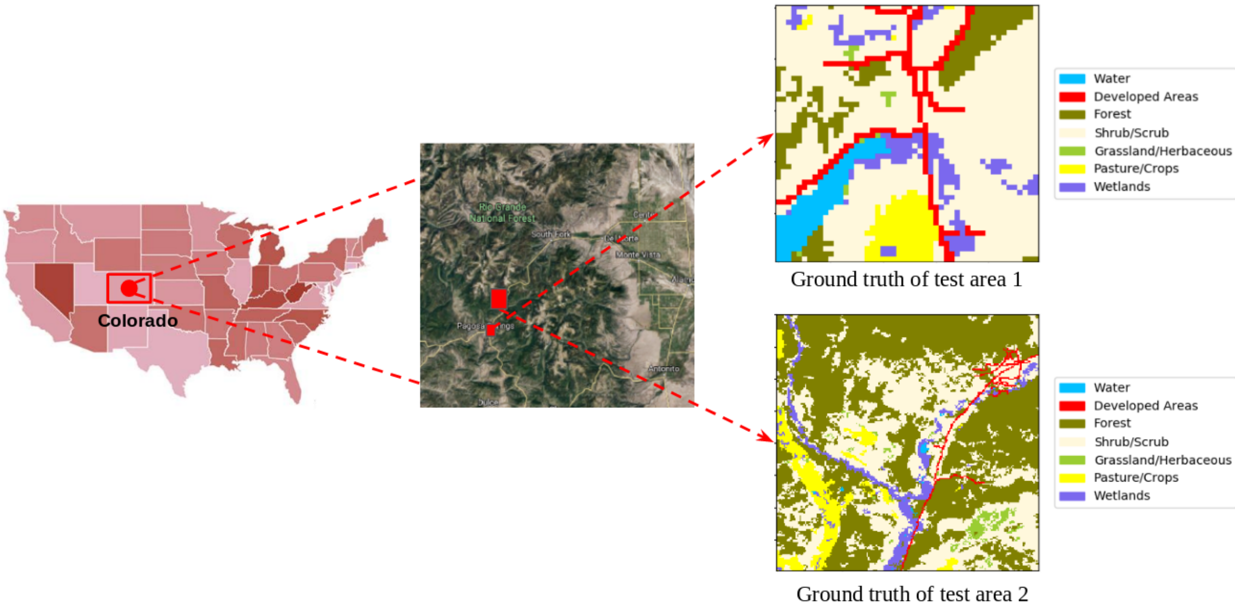


Figure 9: Ground truth of test areas.

7.2.2. Combination of one classification and one clustering methods

In this section, we focus on evaluating the proposed transformation by combining a classification and a clustering. We test nine pairs of combinations between three classifiers: 5-Nearest Neighbors (5-NN), Random Forest (RF) and Stochastic Gradient Boost (SGB) and three clustering methods: K-Means (KM), Spectral Clustering (SC) and Gaussian Mixture Model (GMM). In each combination, we set the number of clusters as 3, 7, 11, 15 in each clustering method to study how this affects the fusion results. We evaluate the results by F1 score with weighted average to take into account the imbalance of labels. Each test was launched 15 times to calculate the average.

Table 11 shows the F1 score for the three fusion methods. We also show the silhouette score of each clustering to measure its quality. It is evident that on the test data set, Karem's and EFSC perform better than EC3 which requires high-quality clustering results. The combination results based on Karem's transformation and EFSC have neither evident relation with the silhouette score of clustering nor the number of partitions. Rather than directly take the agreement of classification and clustering as combination results, these two methods transfer information of clustering into the frame of discernment of classification and fused them by belief functions. During this process, complementary information that is difficult to measure directly is extracted to improve the accuracy of the classification.

Compared with Karem's transformation, the EFSC11 shows no obvious improvement because the combination with clustering can fuse only limited information with the classification. Our proposed transformation is generally more cautious than Karem's, so that it is less possible to change its decisions when the infor-

Table 11: Accuracy of Karem's, EFSC11 and EC3, with combination of one classification (5-NN, RF, SGB) and one clustering (KM, SC,GMM) on test area 1.

Clustering	Partitions (Silhouette score)	5-NN 0.561			RF 0.541			SGB 0.553		
		Karem's	EFSC11	EC3	Karem's	EFSC11	EC3	Karem's	EFSC11	EC3
KM	k=3 (0.424)	0.566	**0.582**	0.568	**0.607**	0.575	0.547	**0.626**	0.604	0.558
	k=7 (0.349)	0.567	**0.568**	0.562	**0.584**	0.561	0.541	**0.629**	0.590	0.553
	k=11 (0.315)	0.570	**0.585**	0.562	**0.586**	0.560	0.542	**0.629**	0.582	0.554
	k=15 (0.294)	0.568	**0.570**	0.562	**0.587**	0.534	0.545	**0.621**	0.570	0.553
SC	k=3 (0.401)	**0.587**	0.584	0.568	**0.587**	0.571	0.552	0.561	**0.567**	0.560
	k=7 (0.119)	**0.579**	0.570	0.563	**0.606**	0.558	0.545	**0.603**	0.579	0.553
	k=11 (0.087)	**0.573**	0.571	0.563	**0.593**	0.572	0.541	**0.600**	0.581	0.553
	k=15 (0.009)	0.570	**0.572**	0.550	0.596	**0.572**	0.541	**0.597**	0.580	0.553
GMM	k=3 (0.239)	0.566	**0.613**	0.564	**0.568**	0.546	0.540	**0.593**	0.558	0.554
	k=7 (0.171)	0.565	**0.575**	0.568	**0.603**	0.556	0.543	**0.618**	0.566	0.554
	k=11 (0.111)	0.559	**0.564**	0.554	**0.632**	0.578	0.542	**0.620**	0.562	0.553
	k=15 (0.155)	0.570	**0.581**	0.558	**0.589**	0.551	0.542	0.570	**0.613**	0.553

mation provided by the clustering is highly limited.

7.2.3. Combination of one classification and multiple clustering methods

In this section, we evaluate the EFSC1m configuration that combines one classification and multiple clustering methods through the proposed iterative fusion process. We select three clustering methods: K-Means (KM), Spectral Clustering (SC) and Gaussian Mixture Model (GMM), which are commonly used in land cover classification, to construct the pool of clustering methods. The number of clusters in each clustering method varies from 3 to 15. We perform the iterative fusion process 300 times by randomly selecting a clustering in the pool.

Since K-Means and SC are implemented based on distance, we employ the distance model to construct their BBAs, as illustrated in section 3.4. Gaussian Mixture Model separates data by distribution, thus providing a probability for the clustering results. The proposed transformation requires separable BBAs from clustering, so the probability from GMM is discounted by a global discounting coefficient 0.9. They can be decomposed by the canonical decomposition to a group of SBBAs, as illustrated in section 3.7.

We have verified that the EFSC11 configuration does not depend on the type of classifiers in section 7.2.2. Due to space limitations, we therefore only displayed the results of Random Forest as a supervised method in the EFSC1m configuration. Karem's transformation only works on a supervised method and a classification method. To compare with the EFSC1m configuration, we also used the proposed iterative fusion process to combine several clustering methods with Karem's transformation. As EC3 cannot handle uncertainty and cooperate with the iterative fusion process, we used EC3 to combine directly the classifier with all clustering methods in the clustering pool.

We display the comparison of accuracy of the three methods in Table 12 and mark the best performance on each class by bold text. EC3 improves the accuracy by 0.8% compared to the original classification, which is outperformed by EFSC1m and Karem's methods. The results of the EC3 method were obviously limited by insufficient training samples. Compared to the results in section 7.2.2, the combination with multiple clustering methods does not make a big difference, as the clustering results on our test data have low qualities as measured by the silhouette score. EC3 generally requires sufficient training samples for parameter selection and good clustering results to reach the agreement.

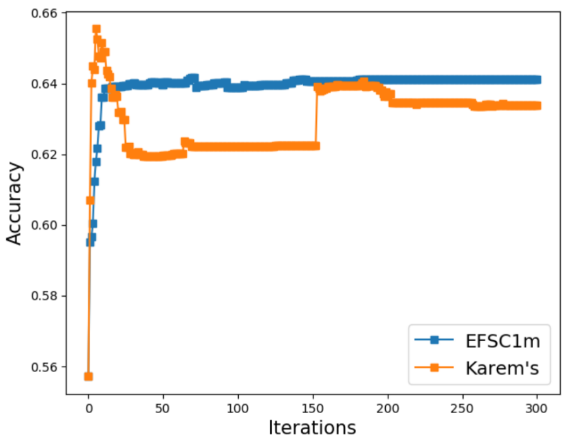
Although Karem's can improve accuracy by 8%, it cannot reserve information on all classes during the iterative fusion process. Some classes, such as *Developed Areas*, are more uncertain so that could be ignored when making decisions based on Karem's transformation. EFSC1m shows a noticeable improvement in accuracy of 9% and the information on all classes is well preserved. Indeed, the Karem's transformation only takes into account the uncertainties of the clustering results while ignoring their imprecision. Moreover, the similarity to achieve their transformation is the proportion of each class in a cluster, which also ignores more information compared to Jaccard Index. Therefore, Karem's transformation tends to make a more certain class become more dominant in the iterative fusion process. On the contrary, our proposed transformation can preserve both uncertainty and imprecision in clustering results, based on which decisions become more cautious. When supervised results are unreliable, it is more appropriate to gradually fuse the clustering information in a more cautious manner. The tests are launched 15 times to calculate the averages. As every pixel can be regarded as an independent result, we consider approximately that each test is run 2500×15 to calculate the confidence interval. The confidence interval with confidence level of95% of EFSC1m is $[0.628, 0.637]$, Karem's $[0.617, 0.626]$, EC3 $[0.543, 0.554]$. The experiments show that EFSC1m is significantly better than Karem's and EC3.

Table 12: Accuracy of the original RF, Karem's EFSC1m and EC3 on test area 1.

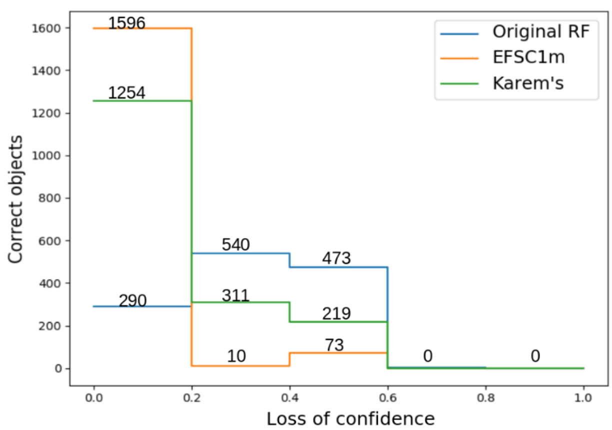
Classifiers	States	Accuracy per class							Accuracy
		Water	Developed Areas	Forest	Shrub/ Scrub	Grassland/ Herbacrous	Pasture/ Crops	Wetlands	
RF	Original	**0.922**	0.277	0.518	0.609	0.095	0.214	**0.215**	0.541
	Karem's	0.795	0.012	0.316	**0.816**	0.056	0.079	0.164	0.622
	EFSC1m	0.866	**0.279**	**0.589**	0.762	**0.109**	**0.335**	0.071	**0.633**
	EC3	0.882	0.263	0.567	0.670	0.042	0.264	0.108	0.549

As EC3 neither works in the iterative process nor deals with uncertainty, we only study the EFSC1m and

Karem's on the changes of accuracy at each iteration and the reduction of uncertainty. Due to the limited space, we display only the results in one experiment in Figure 10. Figure 10a shows how accuracy changes with each iteration. The accuracy of Karem's could decrease during the iterative fusion process while our method become more stable and can reach convergence for a fixed clustering pool. Dempster's combination rule can strengthen similar information so that in the proposed iterative fusion process, it can gradually decrease ignorance. Therefore, in our method, the later a grouping is combined, the less important it is, until it no longer influences the fusion results. We also illustrate the change of loss of confidence before and after fusion in Figure 10b, objects correctly classified in different intervals of loss of confidence are counted. For the original RF, the initial loss of confidence of the correct objects distributes similarly in intervals: $[0., 0.2]$, $(0, 2, 0.4]$, $(0.4, 0.6]$. EFSC1m and Karem's methods can highly reduce the loss of confidence and make the majority of correct objects in the interval $[0, 0.2]$. Compared to the Karem method, EFSC1m has more advantages in making the correct object less uncertain.



(a) Accuracy change in the iterative process



(b) Correct objects in different intervals of loss of confidence

Figure 10: Accuracy change and correct objects in different intervals of loss of confidence of EFSC1m and Karem's on test area 1.

7.2.4. Robustness of the EFSC1m on mislabeled training samples

To further evaluate the effectiveness and robustness of the proposed method, we test it on mislabeled training samples. This is a common problem in land cover classification, as labels and data are generally collected at different times. In this section, we start with a real case where labels were collected in 2011. However, the remote sensing data were collected in 2018 with very different surface circumstances. We evaluate the results on the latest NLCD labels published in 2016, which are quite similar to the reality in the field in 2018. Later, we manually modified the proportions of mislabeled training samples in each class by a uniform distribution to study the evolution of the results. We use the same method configuration as in section 7.2.3. The training sample configuration is detailed in Table 13 in which we illustrate the training samples by class and show the proportion of mislabeled samples in the NLCD 2011.

We employ the NLCD 2011 with 36.8% incorrect labels as training samples for RF and showed the results in Table 14 where the best performance of each class is marked in bold. Karem's and EFSC1m still outperform EC3 with mislabeled training samples. EFSC1m can improve the overall accuracy by 13% than the original RF, which is more satisfying than Karem's and EC3. Due to 97.9% incorrect labels in *Grassland/Herbaceous*, its accuracy after fusion becomes 0 for Karem's and EFSC1m, but for EC3, the accuracy is kept as the original supervised results. For *Pasture/Crops* which has 61.1% incorrect labels, EFSC1m can still improve its accuracy by 6%.

Table 13: Descriptions of training samples in the real case (M: proportion of mislabeled training samples) on test area 1.

	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	Total
Training samples	32	257	135	239	47	36	82	828
M in NLCD 2011	0.218	0.377	0.148	0.230	0.979	0.611	0.707	0.368

Table 14: Accuracy with the mislabeled training samples in the real case on test area 1.

Methods	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	Accuracy
RF	**0.926**	0.234	0.468	0.524	**0.019**	0.301	**0.102**	0.473
Karem's	0.910	0.058	0.184	0.803	0.005	0.012	0.009	0.578
EFSC1m	0.893	**0.273**	**0.539**	**0.713**	0.008	**0.368**	0.079	**0.602**
EC3	**0.926**	0.235	0.493	0.535	**0.019**	0.289	0.100	0.478

In the artificial case, we set the proportion of mislabeled training samples per class from 0.0 to 0.9, as shown in Figure 11. The accuracy of the original RF decreases slightly when $M \in [0, 0.3]$, where Karem's and EFSC1m have similar performance. When $M > 0.5$, the accuracy of Karem' method has obviously decreased but EFSC1m retains the improvement. Compared to Karem's, EFSC1m can improve the overall accuracy by 20% at most. The results of EFSC1m decrease when $M > 0.7$, but they are still more satisfactory than Karem's. Apparently, the cautiousness of the EFSC1m makes it more robust in case the training samples are partially incorrect. EC3 has similar results to the original RF, which is less satisfactory than the other two methods initially. When $M > 0.7$, the accuracy of Karem's and EFSC1m becomes poorer than that of the original RF, whereas EC3 can keep the same. This can be explained by the fact that EC3 is based on consensus maximization, and therefore relies more on classification than clustering. We retrieve classical results obtained in information fusion.

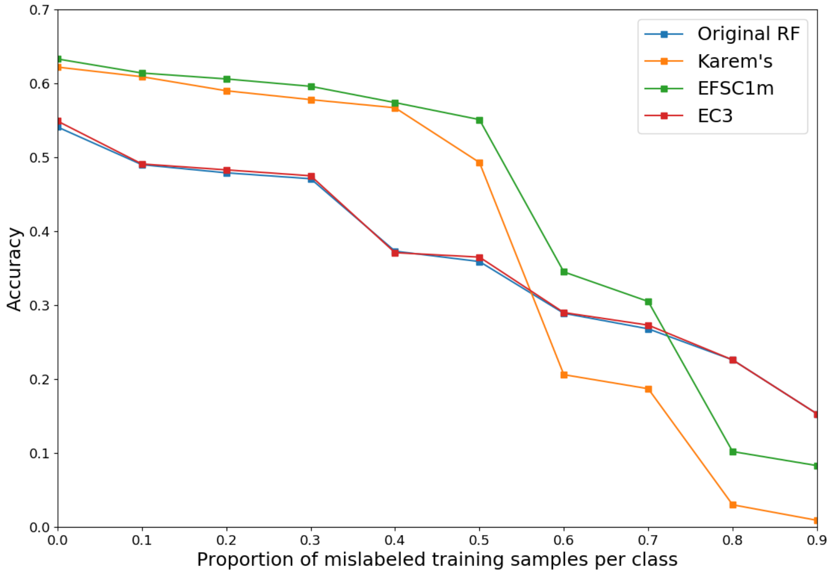
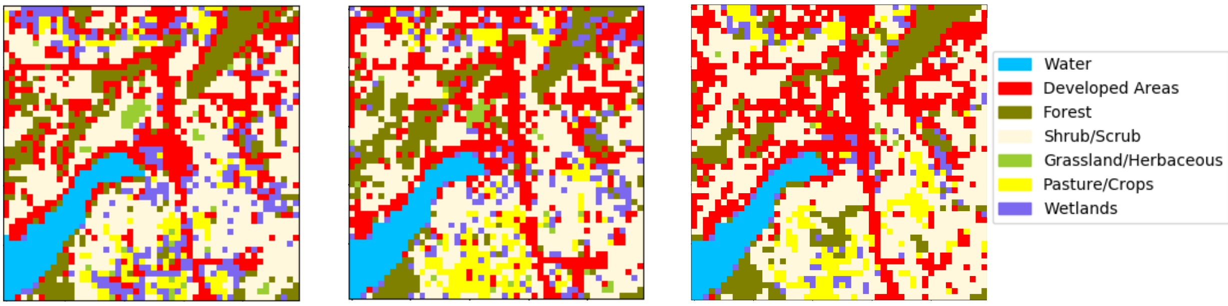


Figure 11: Accuracy with different proportions of mislabeled training samples per class in the artificial case on test area 1.

7.2.5. Combination of multiple classification and clustering methods

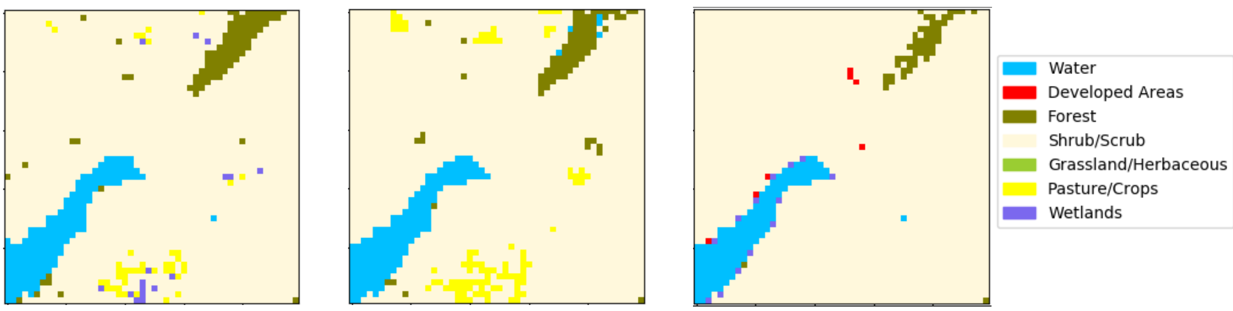
In this section, we evaluate the EFSCmm configuration and compare results with Karem's, EC3, and also the combination of the three classifications by MV. We retake the three supervised classifiers: 5NN, RF, and SGB in section 7.2.2. For each individual classifier, we combine them with multiple clustering methods by EFSC1m configuration. In this way, the reliable information from each classifier is thus reinforced. Information extracted by fusion with multiple clustering methods, from different classification results, is discounted because their ignorance in BBAs usually is near to 0. In the experiment, we set this discounting coefficient as 0.8. Note that this coefficient is only used to fuse information by Dempster's rule. We apply the same configuration for Karem's as EFSCmm. As for EC3, we use it directly to combine the three classifications with all clustering methods. Due to the limitation of space, we do not discuss the details of EFSC1m with 5-NN and SGB. Furthermore, we have already thoroughly studied EFSC1m with RF in the two previous experiments, which can indicate the effectiveness of the EFSC1m using 5-NN or SGB.

Figure 12 shows the results of the original classification of 5-NN, RF, and SGB in one experiment. Note that *Shrub/Scrub* are easy to classify as *Developed Areas*, and *Wetlands* is difficult to identify. For each classification results, we use EFSC1m to combine it with multiple clustering methods. Figure 13 displays Karem's with each classification. Many land covers such as *Developed Areas, Grassland/Herbaceous* and *Wetlands*, are almost eliminated in the fusion results, and *Shrub/Scrub* becomes the dominant class. Apparently, Karem's transformation cannot preserve the weak information during an iterative process where the clustering results may be considered several times. Results of EC3 combining each classification with multiple clustering methods are presented in Figure 14 and show less improvement compared to the results of each original classification. We display the results of EFSC1m with 5-NN, RF, and SGB in Figure 15. The identification of *Developed Areas* is evidently improved and weak information such as *Wetlands* can also be well preserved after fusion with multiple clustering methods.



(a) 5-NN (b) RF (c) SGB

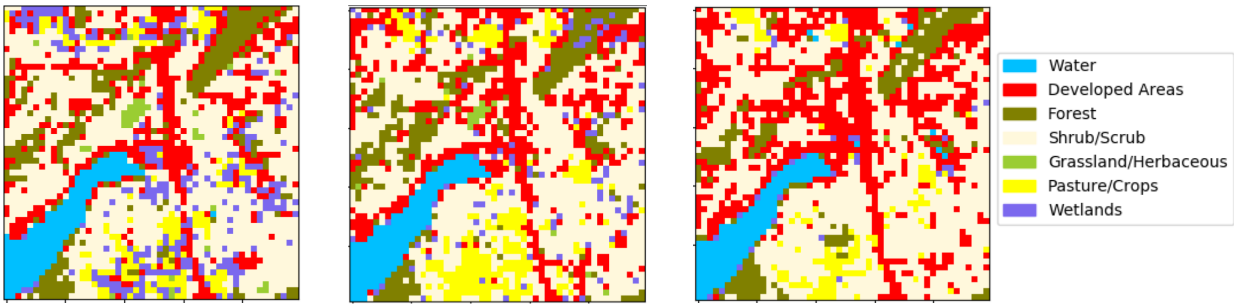
Figure 12: Original classification results of 5-NN, RF and SGB on test area 1.



(a) Karem's with 5-NN (b) Karem's with RF (c) Karem's with SGB

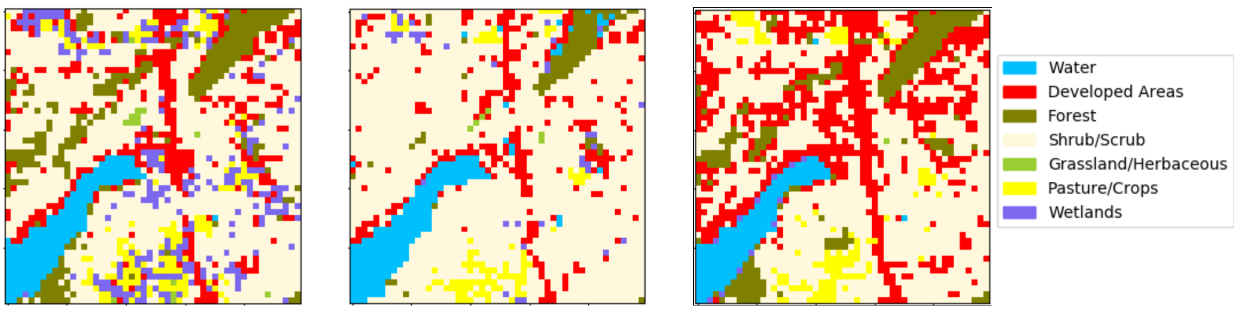
Figure 13: Karem's with combination of one classification (5-NN, RF and SGB) and multiple clustering methods on test area 1 on test area 1.

Results of combination with the three classification results and multiple clustering methods are detailed in Table 15 and Figure 16. The original accuracy of each classifier is also detailed to compare. EC3 and MV have similar performances because the initial classification results are less reliable and the training samples are not sufficient. Both the EFSCmm and Karem's methods can effectively improve accuracy by combining several supervised and unsupervised methods. Karem's, which apparently is less effective in distinguishing all possible classes because its transformation of the BBAs only deals with uncertainty while ignoring imprecision. The transformation we propose carefully takes into account the uncertainty and imprecision in clustering. Therefore, compared to the other two methods, the advantage of the EFSC is to strengthen reliable information and also to prudently reserve unreliable information when merging with multiple clustering methods. The cautiousness of the EFSC makes it more effective when training samples are limited or with incorrect labels. We also calculate the confidence interval with confidence level of 95% and obtain $[0.650, 0.657]$ for EFSCmm, $[0.627, 0.634]$ for Karem's, $[0.557, 0.568]$ for EC3, and $[0.556, 0.567]$ for MV. EFSCmm is significantly better than other methods, and EC3 has similar performance as MV when classification and clustering results have poor qualities.



(a) EC3 with 5-NN (b) EC3 with RF (c) EC3 with SGB

Figure 14: EC3 with combination of one classification (5-NN, RF and SGB) and multiple clustering methods on test area 1.

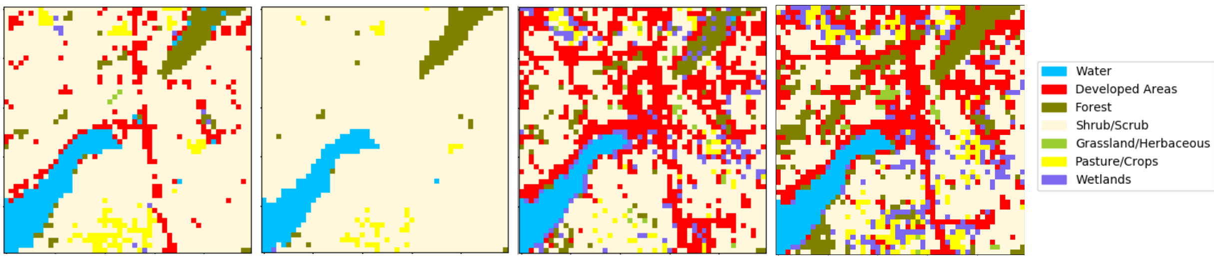


(a) EFSC1m with 5-NN (b) EFSC1m with RF (c) EFSC1m with SGB

Figure 15: EFSC1m with combination of one classification (5-NN, RF and SGB) and multiple clustering methods on test area 1.

Table 15: Accuracy of the original classifications (5-NN, RF, SGB) and four fusion methods (Karem's, EFSCmm, EC3, MV on test area 1.

Methods	Accuracy per class							Accuracy
	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	
5-NN	0.903	0.250	0.558	0.639	**0.142**	0.293	0.188	0.561
RF	0.922	0.277	0.518	0.609	0.095	0.214	**0.215**	0.541
SGB	**0.936**	0.240	0.557	0.635	0.017	0.277	0.133	0.553
Karem's	0.893	0.009	0.494	**0.818**	0.025	0.273	0.014	0.631
EFSCmm	0.854	**0.290**	0.525	0.801	0.111	**0.424**	0.030	**0.654**
EC3	0.917	0.267	0.475	0.654	0.112	0.122	0.192	0.563
MV	0.920	0.182	**0.597**	0.613	0.103	0.287	0.173	0.562



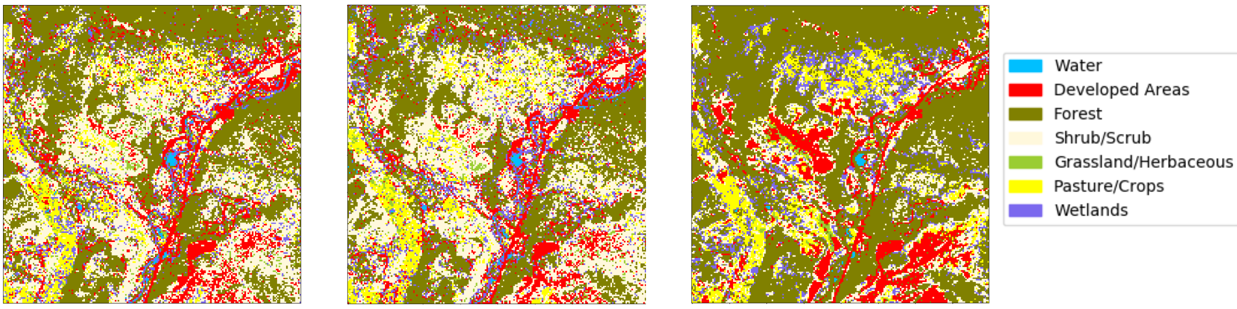
(a) EFSCmm (b) Karem's (c) EC3 (d) MV

Figure 16: EFSCmm, Karem's and EC3 with combination of multiple classification and clustering methods on test area 1.

We also evaluate the proposed EFSC on test area 2 which is 16 times larger than test area 1. The details and ground truth on test area 2 are shown in Table 10 and Figure 9. Due to the limitation of space, we only display the combination results with multiple classification and clustering by EFSCmm, Karem's and EC3. We also show the combination only for classifications by MV. The experiments are launched 5 times on test area 2 to calculate the averages.

The accuracy measured by F1 score for each method are detailed in Table 16. The classification results are shown in Figure 17 and the combination results in Figure 18. Test area 2 is dominant by *Forest* and *Shrub/Scrub*, and classes such as *Water* and *Developed Areas* are more unbalanced compared to teat area 1. EFSCmm improves the accuracy by 7% at most and also outperforms other methods. The confidence interval with confidence level of 95% of EFSCmm is $[0.663, 0.668]$, Karem's $[0.653, 0.658]$, EC3 $[0.630, 0.635]$ and MV $[0.625, 0.631]$. The experiments show that EFSCmm is also significantly better than Karem's, and

EC3 on test area 2.

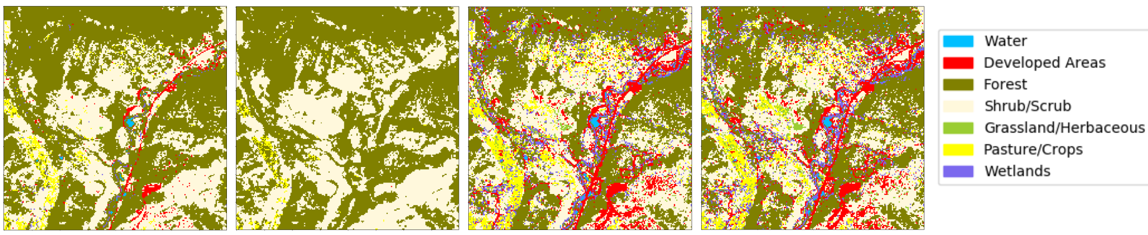


(a) 5NN (b) RF (c) SGB

Figure 17: Original classification results of 5-NN, RF and SGB on test area 2

Table 16: Accuracy of the original classifications (5-NN, RF, SGB) and four fusion methods (Karem's, EFSCmm, EC3, MV) on test area 2.

Methods	Accuracy per class							Accuracy
	Water	Developed Area	Forest	Shrub /Scrub	Grassland /Herbaceous	Pasture /Crops	Wetlands	
5-NN	0.280	0.122	0.733	0.522	0.118	0.379	0.161	0.597
RF	0.245	0.137	0.714	0.553	0.115	0.422	0.185	0.602
SGB	0.359	0.141	0.753	0.483	0.079	0.367	0.183	0.595
Karem's	0.	0.	0.784	**0.660**	0.	0.303	0.	0.656
EFSCmm	**0.431**	**0.227**	**0.798**	0.639	0.016	0.325	0.089	**0.666**
EC3	0.330	0.150	0.756	0.576	0.089	**0.434**	0.195	0.633
MV	0.314	0.143	0.759	0.558	**0.121**	0.428	**0.202**	0.628



(a) EFSCmm (b) Karem's (c) EC3 (d) MV

Figure 18: Combination of multiple classification and clustering by EFSCmm, Karem's, and EC3, and combination of classifications by MV on test area 2.

One of the benefits of belief functions based methods is to correct the information in classifications by clustering. For the three classification results, they are all misclassified parts of *Shrub/Scrub* as *Pasture/Crops*, which is difficult to correct by EC3 and MV. This is because for EC3 only classification results can provide semantic labels so that it depends more on classification than clustering. Belief functions based methods, such as EFSC and Karem's, can transform clustering into the same frame as classification, consequently generating more information on semantic labels to combine.

8. Conclusion

In this paper, we have focused on the fusion of heterogeneous information at the output level. The major difficulty of the combination is to transfer heterogeneous information from supervised and clustering methods into the same scenarios. Under the framework of belief functions, we proposed a transformation to transfer heterogeneous BBAs to the same frame of discernment. Based on this transformation, an iterative strategy was proposed to combine supervised methods with multiple clustering methods.

The proposed method EFSC can efficiently improve the classification accuracy when the information at the raw data level is limited (*i.e.* raw data is inaccessible, training samples are insufficient or partially incorrect). Combining supervised methods with clustering results helps to reduce the uncertainty of supervised results and to enhance reliable information. EFSC emphasizes the treatment of uncertain information in order to make prudent decisions. Different configurations of EFSC have been evaluated both on synthetic data and real remote sensing data, in comparison with two other methods: Karem 's and EC3. EFSC has shown significantly better performance compared to Karem's and EC3. EFSC does not rely heavily on training samples, so it is more pertinent to improve the accuracy of the classification at the output level.

The combination of heterogeneous information at the output level is always difficult. We will further study how supervised and unsupervised results affect the combination in the proposed transformation, especially on regrouping/refining the original classes based on clustering results. In our current approach, we use the Jaccard index as the similarity between classes and clusters. We also consider to study different measurements of similarity and try to understand how the similarity affects the process of the proposed transformation. In our future work, we will focus on the feasibility to generate the similarity matrix by learning process with limited training samples. We will also develop a parallel implementation to test the proposed framework on very large datasets with different types of data.

References

[1] A. Acharya, E. R. Hruschka, J. Ghosh, and S. Acharyya. C3E: A framework for combining ensembles of classifiers and clusterers. In C. Sansone, J. Kittler, and F. Roli, editors, *Multiple Classifier Systems*, pages 269–278, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[2] X. Ao, P. Luo, X. Ma, F. Zhuang, Q. He, Z. Shi, and Z. Shen. Combining supervised and unsupervised models via unconstrained probabilistic embedding. *Information Sciences*, 257:101 – 114, 2014.

[3] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L. He. Fuzziness based semi-supervised learning approach for intrusion detection system. *Information Sciences*, 378:484 – 497, 2017.

[4] Y. Bi, J. Guan, and D. Bell. The combination of multiple classifiers using an evidential reasoning approach. *Artificial Intelligence*, 172(15):1731 – 1751, 2008.

[5] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, pages 92–100, New York, NY, USA, 1998. ACM.

[6] T. Chakraborty. Ec3: Combining clustering and classification for ensemble learning. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 781–786, Nov 2017.

[7] A. Cornuéjols, C. Wemmert, P. Ganarski, and Y. Bennani. Collaborative clustering: Why, when, what and how. *Information Fusion*, 39:81 – 95, 2018.

[8] A. P. Dempster. *A Generalization of Bayesian Inference*, pages 73–104. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[9] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(5):804–813, May 1995.

[10] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2):234 – 264, 2008.

[11] B. E. Erdogan, S. Özöğür Akyüz, and P. K. Ata. A novel approach for panel data: An ensemble of weighted functional margin svm models. *Information Sciences*, 2019.

[12] A. Essaid, A. Martin, G. Smits, and B. Ben Yaghlane. A Distance-Based Decision in the Credal Level. In *International Conference on Artificial Intelligence and Symbolic Computation (AISC 2014)*, pages 147 – 156, Sevilla, Spain, Dec. 2014.

[13] G. Forestier and C. Wemmert. Semi-supervised learning using multiple clusterings with limited labeled data. *Information Sciences*, 361-362:48 – 65, 2016.

[14] D. S. Frossyniotis, M. Pertselakis, and A. Stafylopatis. A multi-clustering fusion algorithm. In *SETN*, 2002.

[15] J. Gao, F. Liang, W. Fan, Y. Sun, and J. Han. Graph-based consensus maximization among multiple supervised and unsupervised models. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 585–593. Curran Associates, Inc., 2009.

[16] J. Gao, F. Liang, W. Fan, Y. Sun, and J. Han. A graph-based consensus maximization approach for combining multiple supervised and unsupervised models. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):15–28, Jan 2013.

[17] Z.-S. Hou and Z. Wang. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3 – 35, 2013. Data-based Control, Decision, Scheduling and Fault Diagnostics.

[18] A.-L. Jousselme, D. Grenier, and E. Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2(2):91 – 101, 2001.

[19] F. Karem, M. Dhibi, and A. Martin. Combination of supervised and unsupervised classification using the theory of belief functions. In T. Denœux and M.-H. Masson, editors, *Belief Functions: Theory and Applications*, pages 85–92, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[20] S. Le Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar. Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):1018–1031, July 1997.

[21] F. Li, Y. Qian, J. Wang, and J. Liang. Multigranulation information fusion: A dempster-shafer evidence theory-based clustering ensemble method. *Information Sciences*, 378:389 – 409, 2017.

[22] N. Li, A. Martin, and R. Estival. An automatic water detection approach based on dempster-shafer theory for multi-

spectral images. In *2017 20th International Conference on Information Fusion (Fusion)*, pages 1–8, 2017.

[23] N. Li, A. Martin, and R. Estival. Combination of supervised learning and unsupervised learning based on object association for land cover classification. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8, 2018.

[24] N. Li, A. Martin, and R. Estival. An automatic land covers identification based on Dempster-Shafer theory for multi-spectral images. In *IGARSS2019*, Yokohama, Japan, July 2019.

[25] Z. Liu, Q. Pan, J. Dezert, J. Han, and Y. He. Classifier fusion with contextual reliability evaluation. *IEEE Transactions on Cybernetics*, 48(5):1605–1618, May 2018.

[26] A. Martin. Comparative study of information fusion methods for sonar images classification. In *2005 7th International Conference on Information Fusion*, volume 2, pages 7, July 2005.

[27] A. Martin, G. Sevellec, and I. Leblond. Characteristics vs decision fusion for sea-bottom characterization. *Colloque Caractérisation in-situ des fonds marins*, Brest, France, 21-22 October 2004.

[28] E. Morvant, A. Habrard, and S. Ayache. Majority vote of diverse classifiers for late fusion. In P. Fränti, G. Brown, M. Loog, F. Escolano, and M. Pelillo, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 153–162, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[29] M. J. A. Patwary and X.-Z. Wang. Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning. *Information Sciences*, 490:93 – 112, 2019.

[30] M. J. A. Patwary, X. Z Wang, and D. Yan. Impact of fuzziness measures on the performance of semi-supervised learning. *International Journal of Fuzzy Systems*, pages 1–13, 2019.

[31] S. Razi, M. R. K. Mollaei, and J. Ghasemi. A novel method for classification of bci multi-class motor imagery task based on Dempster-Shafer theory. *Information Sciences*, 484:14 – 26, 2019.

[32] M. Sabzevari, G. Martínez-Muñoz, and A. Suárez. Vote-boosting ensembles. *Pattern Recognition*, 83:119 – 133, 2018.

[33] F. Sebbak, F. Benhammadi, M. Mataoui, S. Bouznad, and Y. Amirat. An alternative combination rule for evidential reasoning. In *17th International Conference on Information Fusion (FUSION)*, pages 1–8, July 2014.

[34] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[35] P. Smets. Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1 – 35, 1993.

[36] J. Sublime, B. Matei, G. Cabanes, N. Grozavu, Y. Bennani, and A. Cornujols. Entropy based probabilistic collaborative clustering. *Pattern Recognition*, 72:144 – 157, 2017.

[37] J. Zhao, X. Xie, X. Xu, and S. Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43 – 54, 2017.